



Dean, J., Ganesh, A., & Crane, E. (2020). Functional large deviations for Cox processes and Cox/G/ $\infty$  queues, with a biological application. *Annals of Applied Probability*, 30(5), 2465-2490.  
<https://doi.org/10.1214/20-AAP1563>

Peer reviewed version

Link to published version (if available):  
[10.1214/20-AAP1563](https://doi.org/10.1214/20-AAP1563)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Institute of Mathematical Statistics at <https://projecteuclid.org/euclid.aoap/1600157080> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Functional Large Deviations for Cox Processes and $Cox/G/\infty$ Queues, with a Biological Application

Justin Dean, Ayalvadi Ganesh, Edward Crane

November 14, 2019

## Abstract

We consider an infinite-server queue into which customers arrive according to a Cox process and have independent service times with a general distribution. We prove a functional large deviations principle for the equilibrium queue length process. The model is motivated by a linear feed-forward gene regulatory network, in which the rate of protein synthesis is modulated by the number of RNA molecules present in a cell. The system can be modelled as a non-standard tandem of infinite-server queues, in which the number of customers present in a queue modulates the arrival rate into the next queue in the tandem. We establish large deviation principles for this queueing system in the asymptotic regime in which the arrival process is sped up, while the service process is not scaled.

# 1 Introduction

The work in this paper is motivated by the problem of modelling fluctuations in the number of protein molecules in a cell. The synthesis of proteins is catalysed by RNA molecules, which in turn are transcribed from DNA molecules. Both RNA and protein molecules degrade spontaneously after some random time. It is important for proper functioning of the cell that protein numbers are maintained within certain limits, and biologists are interested in understanding the regulatory mechanisms involved in controlling their fluctuations. Consequently, the problem of modelling stochastic fluctuations has attracted interest, and there has been considerable work on Markovian models of such systems; see, e.g., [11, 13]. These models assume that each copy of a gene creates RNA molecules according to a Poisson process (while active), that each RNA molecule generates protein molecules according to a Poisson process, and that the lifetimes of RNA and protein molecules are exponentially distributed. The assumption of exponential lifetimes is biologically unrealistic; for example, inhomogeneities in the cellular environment could result in lifetimes that are mixtures of exponential distributions, or the denaturing of molecules could be a multistage process.

Our approach relies on modelling the chemical kinetics using  $\cdot/G/\infty$  queues rather than Markov processes, which correspond to  $\cdot/M/\infty$  queues. Customer arrivals into the queue correspond to the synthesis of molecules of a specified type; after independent lifetimes with a general distribution, the molecules decay which equates to service (and departure) of the corresponding customers. For the problem described above, we have two such queues in series, one for RNA molecules and one for proteins. However, unlike in a tandem queueing network, where departures from one queue enter the next queue in series, here departures just leave the system; the way influence propagates is that the arrival rate into the protein queue is modulated by the occupancy of the preceding queue (here, RNA) in the series. We consider a very simple form of modulation, in which the arrival rate into a queue is proportional to the occupancy of the preceding queue, and the arrival process is conditionally Poisson given the occupancy. Thus, this results in a Cox process model for the arrivals into a queue, and the system is modelled as a series of  $Cox/G/\infty$  queues interacting as described.

We briefly recall the description of the queue length process in an  $M/G/\infty$  queue with arrival rate  $\lambda$  and service distribution  $F$ . The arrival process into this queue can be represented as an inhomogeneous Poisson process on  $\mathbb{R} \times \mathbb{R}_+$  with intensity measure  $\lambda \otimes F$ . If a realisation of this point process has a point at  $(t, y)$ , it denotes that a customer arrives at time  $t$  bringing

a service requirement of  $y$ . The queue length at time  $t$  is simply the total number of points of the Poisson process in the set

$$A_t = \{(s, y) : s \leq t, y > t - s\},$$

as a customer arriving at time  $s$  will still be in the system at time  $t$  if and only if its service requirement is greater than  $t - s$ . (We follow the convention of defining the queue length process to be right continuous.) Likewise, the queue length process during a time interval  $[s, t]$  can be described in terms of the empirical measure of the above Poisson process on the wedge-shaped set

$$A_{[s, t]} = \bigcup_{u \in [s, t]} A_u.$$

In the problem we want to study, the intensity of the arrival process is modulated by the number of customers present in the previous queue. Hence, we need to model it as a Cox process and study the corresponding  $Cox/G/\infty$  queue. As described above, this requires us to study the empirical measure of a Cox process on a subset of  $\mathbb{R}^2$ . We shall in fact study them in a more general setting of  $\sigma$ -compact Polish spaces, namely Polish spaces that can be covered by countably many compact subsets. Our goal is to obtain functional large deviation principles (FLDPs) for the corresponding queue length processes; we shall obtain these by contraction from LDPs for the empirical measure of the Cox process. We have not been able to drop the technical assumption of  $\sigma$ -compactness from our proof, but do not know if it is essential for the stated results.

In terms of the motivating application, regulatory mechanisms within cells need to be robust enough to filter out small fluctuations, but sensitive to variations of larger magnitude. For example, p53 is a tumour suppressor protein and master transcriptional regulator of response to DNA damage, and the gene that encodes it is the most frequently mutated gene in cancer cells. Normal proliferating cells typically respond to DNA damage with repair mechanisms, or by activating apoptosis (programmed cell death) or senescence (prolonged arrest of the cell cycle). Large fluctuations in p53 concentrations can alter such a cell fate, triggering a switch from an arrested to a proliferative state [14]. Similarly, bistable switches are a very common motif in gene regulatory networks where significant oscillations in protein numbers can result in changes in cell fate [15]. While biologists have long been interested in studying fluctuations in molecule numbers, most work to date has focused on second-order statistics [11]. The scaling regime studied in this paper might be more relevant for understanding the rare but large

fluctuations that trigger changes in cell fate of the type described above. In addition, a functional LDP of the type presented here can be used to identify the typical sample paths leading to rare events of interest, and the typical time scales over which they occur, by solving the optimization problem in the definition of the rate function. The rate functions are explicit enough that the solution of these optimization problems can be well approximated numerically, or upper and lower bounds calculated, in concrete applications.

We present our model and main results in the next section, followed by the proofs in the final two sections.

## 2 Model and Results

We now set out our Cox process model. Let  $(E, d)$  be a  $\sigma$ -compact Polish space, and let  $\Lambda$  be a random finite Borel measure on  $E$ ; in other words,  $\Lambda$  is a random variable taking values in  $\mathcal{M}_+^f(E)$ , the space of finite non-negative Borel measures on  $E$ . A Cox process  $\Phi$  with stochastic intensity  $\Lambda$  is a point process which is conditionally Poisson, with intensity measure  $\lambda$  on the event that  $\Lambda = \lambda$ . Note that the point process  $\Phi$  is almost surely finite. A realisation of  $\Phi$  can be thought of as either a point set  $\{x_1, x_2, \dots, x_k\}$ , or as a counting measure  $\sum_{i=1}^k \delta_{x_i}$ , where  $k$  is the (random) number of points in the realisation. We call the latter the empirical measure corresponding to the realisation of the point set, and note that it is also an element of  $\mathcal{M}_+^f(E)$ . There are three topologies on  $\mathcal{M}_+^f(E)$  which will be of interest to us. We say that a sequence of measures  $\mu_n \in \mathcal{M}_+^f(E)$  converges to  $\mu \in \mathcal{M}_+^f(E)$  in the weak topology if  $\int_E f d\mu_n$  converges to  $\int_E f d\mu$  for all bounded continuous functions  $f : E \rightarrow \mathbb{R}$ ; we say the measures converge in the vague topology if the integrals converge only for continuous functions with compact support (which are necessarily bounded), and that they converge in the tempered topology if the integrals converge for continuous functions with compact support, which furthermore vanish on the boundary of their support. The tempered topology will play a role in Section 4, where we establish an LDP for the queue occupancy measure by expressing it as a function of the empirical measure of the marked arrival process, and using the contraction principle. In order to do so, we require continuity of the queueing map, which holds in the tempered topology on the image space, but not the weak topology.

We now consider a sequence of Cox point processes  $\Phi_n$ , with corresponding stochastic intensities  $\Lambda_n$ . Our first contribution is a large deviation principle (LDP) for their scaled empirical measures:

**Theorem 2.1.** *Suppose that  $(\Lambda_n, n \in \mathbb{N})$  is a sequence of random finite Borel measures on a  $\sigma$ -compact Polish space  $(E, d)$ , and that the sequence  $\Lambda_n/n$  satisfies an LDP in  $\mathcal{M}_+^f(E)$  equipped with the weak topology, with good rate function  $\mathfrak{I}_1(\cdot)$ . Let  $\Phi_n$  be a Cox process with stochastic intensity  $\Lambda_n$ , i.e., a random counting measure on  $E$  equipped with its Borel  $\sigma$ -algebra. Then the sequence of measures  $\Phi_n/n$  satisfies an LDP in  $\mathcal{M}_+^f(E)$  equipped with the weak topology, with good rate function*

$$\mathfrak{I}_2(\mu) = \begin{cases} \inf_{\lambda} \{ \mathfrak{I}_1(\lambda) + \lambda(E) \}, & \text{if } \mu \equiv 0, \\ \inf_{\lambda} \left\{ \mathfrak{I}_1(\lambda) + I_{Poi}(\mu(E), \lambda(E)) + \mu(E) H\left(\frac{\mu}{\mu(E)} \mid \frac{\lambda}{\lambda(E)}\right) \right\}, & \text{if } \mu \neq 0, \end{cases}$$

where  $H$  and  $I_{Poi}$  are defined as follows:

$$H(\beta|\alpha) = \begin{cases} \int \log(d\beta/d\alpha) d\beta & \text{if } \beta \ll \alpha \text{ and } \int |\log(d\beta/d\alpha)| d\beta < \infty \\ +\infty & \text{otherwise,} \end{cases}$$

$$I_{Poi}(x, \alpha) = \begin{cases} x \log \frac{x}{\alpha} - x + \alpha, & \text{if } \alpha > 0, \\ 0, & \text{if } \alpha = 0, x = 0, \\ +\infty, & \text{if } \alpha = 0, x > 0. \end{cases}$$

The function  $H(\beta|\alpha)$  is called the relative entropy or Kullback-Leibler divergence of  $\beta$  with respect to  $\alpha$ .

A slightly different version of this theorem, with only local finiteness of the measures  $\Lambda_n$  assumed, has been established by Schreiber [16], albeit in the vague rather than the weak topology; his result also requires a technical assumption about the measures  $\Lambda_n/n$  dominating a fixed measure with full support on  $E$ , which we do not need. However, his result does not require that the space be  $\sigma$ -compact. The extension of the result to the weak topology is non-trivial, and relies on the finiteness assumption on the intensity measures. In addition, our proof techniques are very different. A functional LDP for rescaled Poisson random measures is proved in [7] using projective limits, and in [10] using Cramér's theorem and subadditivity arguments.

The claim of Theorem 2.1 appears intuitive from the assumed LDP for the intensity measures  $\Lambda_n/n$ , the LDP for a Poisson random variable, and Sanov's theorem for the empirical distribution. However, a number of technical conditions need to be checked. Moreover, while these imply an LDP, goodness of the rate function is not immediate. We show this indirectly by establishing exponential tightness; this is the step where finiteness of the measures is crucial.

Next, we consider a sequence of stationary  $Cox/G/\infty$  queues where the arrival processes are sped up by the index  $n \in \mathbb{N}$ , while the service process

remains unchanged. More precisely, the service times are iid with some fixed distribution  $F$  that does not depend on  $n$ , while the arrival process into the  $n^{\text{th}}$  queue is a Cox process with stochastic intensity (directing measure)  $\Lambda_n$  on  $\mathbb{R}$ . We make the following assumptions.

**Assumptions**

A1  $(\Lambda_n, n \in \mathbb{N})$  is a sequence of random  $\sigma$ -finite measures on  $\mathbb{R}$ , whose laws are translation invariant, such that  $\mathbb{E}[\Lambda_n([a, b])] = n\lambda(b - a)$ , for some fixed  $\lambda > 0$ , and any compact interval  $[a, b] \subset \mathbb{R}$ .

A2 For any interval  $[a, b]$ , the sequence  $(\Lambda_n/n)|_{[a, b]}$  obeys an LDP on  $\mathcal{M}_+^f([a, b])$  equipped with the weak topology, with good rate function  $I_{[a, b]}$ .

A3 Define

$$\psi_n(\theta) = \log \mathbb{E} \left[ e^{\frac{\theta \Lambda_n([0, 1])}{n}} \right].$$

There is a neighbourhood of 0 on which  $\psi_n(n\theta)/n$  is bounded, uniformly in  $n$ .

A4 The mean service time, given by  $\int_0^\infty x dF(x) = \int_0^\infty \bar{F}(x) dx$ , is finite; here  $\bar{F} = 1 - F$  denotes the complementary cumulative distribution function of the service time.

Let  $Q_n(t)$  denote the number of customers at time  $t$  in the infinite-server queue with Cox process arrivals with intensity  $\Lambda_n$  and iid service times with distribution  $F$ . Let  $L_n$  denote the measure on  $\mathbb{R}$  which is absolutely continuous with respect to Lebesgue measure, with density  $Q_n(\cdot)$ . Our second contribution in this paper is the following:

**Theorem 2.2.** *Consider a sequence of Cox/G/ $\infty$  queues indexed by  $n \in \mathbb{N}$ , where the arrival process into the  $n^{\text{th}}$  queue is a Cox process with directing measure  $\Lambda_n$ , and service times are iid with common distribution  $F$ . Suppose the arrival and service processes satisfy Assumptions [A1]-[A4]. Let  $Q_n(t)$  denote the number of customers in the  $n^{\text{th}}$  queue at time  $t$ , and let  $L_n$  denote the random measure on  $\mathbb{R}$  which is absolutely continuous with respect to Lebesgue measure and has density  $Q_n(\cdot)$ . Then the sequence of measures  $L_n$  satisfies Assumptions [A1]-[A3]. In particular, for any compact interval  $[a, b] \subset \mathbb{R}$ , the measures  $(L_n/n)|_{[a, b]}$  satisfy an LDP on  $\mathcal{M}_+^f([a, b])$  equipped with the weak topology, with a good rate function  $J_{[a, b]}$ .*

A fuller description of the rate function  $J_{[a,b]}$  is provided in the proof of this theorem, in Section 4. The theorem shows that the sequence of queue occupancy measures  $L_n$  also satisfies the above assumptions and, in particular, that they satisfy an LDP. This implies that our analysis extends easily to an arbitrary number of  $Cox/G/\infty$  queues in (non-standard) tandem, where the arrivals into each queue constitute a Cox process with directing measure given by the number in the previous queue.

Let  $\Phi_n$  denote the Cox point process of arrivals into the  $n^{\text{th}}$  system as above, with directing measure  $\Lambda_n$ . Denote by  $\hat{\Phi}_n$  the marked point process obtained by marking each arrival with its service time. Can we prove an LDP, not just for the queue occupancy measures but for the queue lengths at a fixed time, say for the sequence of random variables  $Q_n(0)/n = \hat{\Phi}_n(A_0)/n$ ? Unfortunately, the map  $\mu \mapsto \mu(A_0)$  is not continuous in the weak topology, since the indicator function of the set  $A_0$  is not a continuous function. Hence, our approach of invoking the contraction principle does not work. It might be possible to get around this, by sandwiching the indicator function of  $A_0$  between bounded continuous functions which converge to it pointwise from below and above. We could then prove an LDP for the integral of the queue occupancy measure against these functions. If we could calculate the rate function explicitly, and show that it approaches the same limit for the functions approximating the indicator from above and from below, then that would prove the LDP for the marginal queue length distribution. But as these calculations are quite involved, and distract from the main motivation of the present work, we do not pursue them here.

Next, we turn to the departure process from a  $Cox/G/\infty$  queue. While we study non-standard tandems in which departures leave the network rather than feeding the next queue in the tandem, our model can be easily extended to reaction networks in which the products of one reaction are reactants in the next, rather than catalysts as in our model. In that case, one would have a standard tandem of infinite-server queues, instead of the non-standard tandems that are the focus of this paper. In addition, the departure process is an object of interest in queueing theory. With these motivations, we now describe our results for the departure process.

Let  $\Psi_n$  denote the point process of departures, which may be viewed as a random counting measure on  $\mathbb{R}$ . From the description of the  $\cdot/G/\infty$  queue in terms of point processes given in the Introduction, we see that for any interval  $[a, b]$ , we have

$$\Psi_n([a, b]) = \hat{\Phi}_n(\text{cl}(A_{[a,b]} \setminus A_b)), \quad (1)$$

since a customer departs during the interval  $[a, b]$  only if it arrives at time



$t \leq b$ , bringing in an amount of work  $x$  such that  $a \leq t + x \leq b$ ; here  $\text{cl}(B)$  denotes the closure of a subset  $B$  of  $\mathbb{R}^2$ . Our next result establishes an LDP for the empirical measures,  $\Psi_n$ , of the departures from the queue. Hence, the results extend easily to a (standard) tandem of such queues.

**Theorem 2.3.** *Let  $\Phi_n$ ,  $n \in \mathbb{N}$ , be a sequence of Cox arrival processes satisfying Assumptions [A1]-[A3], and let  $\hat{\Phi}_n$  be a Cox process obtained by marking the arrivals with iid service times drawn according to a distribution  $F$  satisfying Assumption [A4]. Let  $\Psi_n$  denote the corresponding departure process from an infinite-server queue, as defined precisely in (1). Then,  $(\Psi_n, n \in \mathbb{N})$  satisfies [A1]-[A3]; in particular, for any fixed compact interval  $[a, b]$ ,  $(\Psi_n/n)|_{[a, b]}$  obeys an LDP on  $\mathcal{M}_+^f([a, b])$  equipped with the weak topology, with a good rate function  $K_{[a, b]}$ .*

The Cox/G/ $\infty$  model studied in this paper is an instance of a queue in a random environment. The first study of infinite-server queues in random environment was in [12]: factorial moments in stationarity were derived for the  $M/M/\infty$  queue in a Markovian environment, namely one in which the arrival and service rates are modulated by a finite state, irreducible, continuous time Markov chain. There has recently been extensive further study of this model, including moments for steady state and transient distributions, and large deviation and central limit asymptotics for the marginal distribution of the queue length; see [2] for a collation of the results. The Markovian assumption on the environment is relaxed in [9], where the background process modulating arrivals and services in an  $M/M/\infty$  queue is just a general càdlàg stochastic process. An LDP is proved for the queue length at an arbitrary fixed time,  $t$ , whereas we establish a process level LDP, without assuming (conditionally) exponential service times. A special type of Cox background process is considered in [8], which proves a functional CLT for the scaled queue length process. In all of these cases the queue length is viewed as a random càdlàg function, whereas we view it as living on a space of measures.

The proof of Theorem 2.1 is presented in Section 3, and the proofs of Theorems 2.2 and 2.3 in Section 4.

### 3 Proof of Empirical Measure LDP

Our proof of Theorem 2.1 relies on a theorem of Chaganty [4], which essentially states that a sequence of probability measures on a product space satisfies an LDP if the corresponding sequences of marginal and conditional

probability distributions do so, and certain additional technical conditions are satisfied. For completeness, we include below a statement of this theorem, together with an extension of Sanov's theorem by Baxter and Jain [1] which is needed to check its conditions, and relevant definitions.

**Definition 3.1.** Let  $(\Omega_1, \mathcal{B}_1)$  and  $(\Omega_2, \mathcal{B}_2)$  be two Polish spaces with their associated Borel  $\sigma$ -fields. Let  $\{\nu_n(\cdot, \cdot)\}$  be a sequence of transition functions on  $\Omega_1 \times \mathcal{B}_2$ , i.e.,  $\nu_n(x_1, \cdot)$  is a probability measure on  $(\Omega_2, \mathcal{B}_2)$  for each  $x_1 \in \Omega_1$  and  $\nu_n(\cdot, B_2)$  is a measurable function on  $\Omega_1$  for each  $B_2 \in \mathcal{B}_2$ . We say that the sequence of probability transition functions  $\{\nu_n(x_1, \cdot), x_1 \in \Omega_1\}$  satisfies the LDP continuously in  $x_1$  with rate function  $J(x_1, x_2)$ , or simply the LDP continuity condition holds, if:

1. For each  $x_1 \in \Omega_1$ ,  $J(x_1, \cdot)$  is a good rate function on  $\Omega_2$ , i.e., it is non-negative, lower semicontinuous (l.s.c.), and has compact level sets.
2. For any sequence  $\{x_{1n}\}$  in  $\Omega_1$  such that  $x_{1n} \rightarrow x_1$ , the sequence of measures  $\{\nu_n(x_{1n}, \cdot)\}$  on  $\Omega_2$  obeys the LDP with rate function  $J(x_1, \cdot)$ .
3.  $J(x_1, x_2)$  is l.s.c. as a function of  $(x_1, x_2)$ .

**Theorem 3.2.** ([4, Theorem 2.3]) Let  $(\Omega_1, \mathcal{B}_1)$ ,  $(\Omega_2, \mathcal{B}_2)$  be two Polish spaces with their associated Borel  $\sigma$ -fields. Let  $\{\mu_{1n}\}$  be a sequence of probability measures on  $(\Omega_1, \mathcal{B}_1)$ . Let  $\{\nu_n(x_1, B_2)\}$  be a sequence of probability transition functions defined on  $\Omega_1 \times \mathcal{B}_2$ . We define the joint distribution  $\mu_n$  on the product space  $\Omega_1 \times \Omega_2$ , and the marginal distribution  $\mu_{2n}$  on  $\Omega_2$  by

$$\mu_n(B_1 \times B_2) = \int_{B_1} \nu_n(x_1, B_2) d\mu_{1n}(x_1), \quad \mu_{2n}(B_2) = \mu_n(\Omega_1 \times B_2).$$

Suppose that the following two conditions are satisfied:

1.  $\{\mu_{1n}\}$  satisfies an LDP with good rate function  $I_1(x_1)$ .
2.  $\{\nu_n(\cdot, \cdot)\}$  satisfies the LDP continuity condition with a rate function  $J(x_1, x_2)$ .

Then the sequence of joint distributions  $\{\mu_n\}$  satisfies a weak LDP on the product space  $\Omega_1 \times \Omega_2$ , with rate function

$$I(x_1, x_2) = I_1(x_1) + J(x_1, x_2).$$

The sequence of marginal distributions  $\mu_{2n}$  satisfies an LDP with rate function

$$I_2(x_2) = \inf_{x_1 \in \Omega_1} [I_1(x_1) + J(x_1, x_2)].$$

Finally,  $\{\mu_n\}$  satisfies the LDP if  $I(x_1, x_2)$  is a good rate function.

**Remark.** Recall that a sequence of probability measures (or random variables) is said to satisfy a weak LDP if the large deviations upper bound holds for all compact sets, and to satisfy a (full) LDP if it holds for all closed sets. For both, the large deviations lower bound holds for all open sets.

**Theorem 3.3.** ([1], Theorem 5) Let  $(S, d)$  be a Polish space. Let  $\{\alpha_n\}$  be a sequence of probability measures on  $(S, d)$  converging weakly to a probability measure  $\alpha$ . For each  $n$ , let  $X_i^n$ ,  $i \in \mathbb{N}$  be iid  $S$ -valued random variables with common distribution  $\alpha_n$ . Let  $\mathcal{M}_1(S)$  denote the space of probability measures on  $S$  and let  $\bar{\mu}_n \in \mathcal{M}_1(S)$  denote the empirical distribution,  $(\delta_{X_1^n} + \dots + \delta_{X_n^n})/n$ . Then  $\{\bar{\mu}_n\}$  satisfies the LDP with good rate function  $H(\cdot|\alpha)$ , which was defined in the statement of Theorem 2.1.

The proof of Theorem 2.1 proceeds through a sequence of lemmas. We begin with an elementary LDP for a sequence of Poisson random variables.

**Lemma 3.4.** Let  $N_n, n \in \mathbb{N}$  be a sequence of Poisson random variables with parameter  $n\alpha_n$ , and suppose that  $\alpha_n$  tends to  $\alpha \geq 0$ . Then the sequence  $N_n/n$  obeys an LDP in  $\mathbb{R}_+$  with good rate function  $I_{Poi}(\cdot, \alpha)$  defined in the statement of Theorem 2.1.

*Proof.* We apply the Gärtner-Ellis theorem [5, Theorem 2.3.6] to the sequence  $N_n/n$ . By direct calculation,

$$\frac{1}{n} \log \mathbb{E} \left[ e^{n\theta \frac{N_n}{n}} \right] = \alpha_n (e^\theta - 1).$$

This sequence of scaled log-moment generating functions converges pointwise to the limit  $\alpha(e^\theta - 1)$ , which is finite and differentiable everywhere (hence also continuous, and essentially smooth). Hence, by the Gärtner-Ellis theorem, the sequence of random variables  $N_n/n$  obeys an LDP with a rate function which is the convex conjugate of  $\alpha(e^\theta - 1)$ . A straightforward calculation confirms that this is the function  $I_{Poi}(\cdot, \lambda)$  in the statement of the lemma, and that it is l.s.c. with compact level sets for each  $\alpha$ .  $\square$

The next two lemmas establish conditional LDPs for the scaled empirical measures of Poisson processes whose scaled intensities converge to a limit.

**Lemma 3.5.** *Let  $\Phi_n, n \in \mathbb{N}$  be a sequence of Poisson point processes with intensity measures  $n\lambda_n \in \mathcal{M}_+^f(E)$ , and suppose that  $\lambda_n$  converge weakly in  $\mathcal{M}_+^f(E)$  to the zero measure. Then,  $\Phi_n/n, n \in \mathbb{N}$  satisfy the LDP in  $\mathcal{M}_+^f(E)$  equipped with the weak topology, with good rate function*

$$\mathcal{I}_0(\mu) = \begin{cases} 0, & \text{if } \mu \equiv 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

*Proof.* As the map  $\mu \mapsto \mu(E)$  is weakly continuous (the indicator of  $E$  is a bounded, continuous function), it follows that  $\lambda_n(E)$  tends to  $\lambda(E) = 0$ . Let  $N_n = \Phi_n(E)$  denote the total number of points in the Poisson process  $\Phi_n$ . Then,  $N_n$  is a Poisson random variable with parameter  $n\lambda_n(E)$ , and it follows from Lemma 3.4 that  $(N_n/n, n \in \mathbb{N})$  obey an LDP with good rate function

$$I_{Poi}(x, 0) = \begin{cases} 0, & \text{if } x = 0, \\ +\infty, & \text{if } x > 0. \end{cases}$$

Let  $F \subset \mathcal{M}_+^f(E)$  be closed in the weak topology, and suppose that it does not contain the zero measure. Define

$$x_F = \inf\{\mu(E) : \mu \in F\}.$$

We claim that  $x_F > 0$ . Indeed, if  $x_F = 0$ , then we can find a sequence of measures  $\mu_n \in F$  such that  $\mu_n(E)$  tends to zero, i.e.,  $\int_E 1 d\mu_n$  tends to zero. It follows that  $\int_E f d\mu_n$  tends to zero for all bounded, measurable, non-negative functions  $f$ , and hence also for all bounded measurable functions. Hence, the sequence  $\mu_n$  converges weakly to the zero measure, contradicting the assumption that  $0 \notin F$  and  $F$  is closed.

We now have the large deviations upper bound for  $F$ :

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n}{n} \in F\right) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n(E)}{n} \geq x_F\right) \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{N_n}{n} \geq x_F\right) = -\infty, \end{aligned}$$

where we have used the LDP for  $N_n/n$  with rate function  $I_{Poi}(\cdot, 0)$  and the fact that  $x_F > 0$  to obtain the last equality.

The large deviations lower bound is trivial for open sets  $G$  not containing the zero measure, as the infimum of the rate function is infinite on such sets. Now, for  $G$  containing the zero measure, we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n}{n} \in G\right) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n}{n} \equiv 0\right) = \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(N_n = 0) \\ &= \liminf_{n \rightarrow \infty} (-\lambda_n(E)) = -\lambda(E) = 0, \end{aligned}$$

as  $N_n \sim \text{Poi}(n\lambda_n(E))$ . This completes the proof of the lemma.  $\square$

**Lemma 3.6.** *Let  $\Phi_n, n \in \mathbb{N}$  be a sequence of Poisson point processes with intensity measures  $n\lambda_n$ , and suppose that the sequence  $\lambda_n$  converges in the weak topology on  $\mathcal{M}_+^f(E)$  to  $\lambda \neq 0$ . Then,  $\Phi_n/n, n \in \mathbb{N}$  satisfy the LDP in  $\mathcal{M}_+^f(E)$  equipped with the weak topology, with good rate function*

$$\mathcal{I}_1(\mu) = \begin{cases} I_{\text{Poi}}(\mu(E), \lambda(E)) + \mu(E)H\left(\frac{\mu}{\mu(E)} \mid \frac{\lambda}{\lambda(E)}\right), & \text{if } \mu \neq 0, \\ I_{\text{Poi}}(0, \lambda(E)), & \text{if } \mu \equiv 0. \end{cases}$$

Here,  $I_{\text{Poi}}(\cdot, \cdot)$  and  $H(\cdot \mid \cdot)$  are as defined in Lemma 3.4 and Theorem 3.3 respectively.

*Proof.* We will prove the lemma by first establishing an LDP for the sequence  $N_n/n$ , then verifying that conditional on this,  $\Phi_n/n$  satisfies the LDP continuously, and invoking Theorem 3.2.

The LDP for  $N_n/n$ , with rate function  $I_{\text{Poi}}(\cdot, \lambda(E))$ , is immediate from Lemma 3.4 since  $\lambda_n(E)$  tends to  $\lambda(E)$ . We now prove an LDP for  $\Phi_n/n$ , conditional on  $N_n/n$ . Fix a sequence  $N_n$  such that  $N_n/n \rightarrow x \geq 0$ . If  $x = 0$ , then the proof follows that of Lemma 3.5, and yields  $\mathcal{I}_0$  as the rate function.

It remains to consider  $x > 0$ . We can write

$$\Phi_n = \delta_{X_1^n} + \delta_{X_2^n} + \dots + \delta_{X_{N_n}^n},$$

where the  $X_i^n$  are iid, with law  $\frac{\lambda_n}{\lambda_n(E)}$ . Note that the probability law of  $X_i^n$  is well-defined for all  $n$  sufficiently large, as  $\lambda_n(E)$  tends to  $\lambda(E) > 0$ . Define

$$\hat{\Phi}_n = \delta_{X_1^n} + \delta_{X_2^n} + \dots + \delta_{X_{[nx]}^n},$$

where the dependence of  $\hat{\Phi}_n$  on  $x$  has been suppressed in the notation. We claim that the sequences  $\Phi_n/n$  and  $\hat{\Phi}_n/n$  are exponentially equivalent (see [5, Definition 4.2.10]). To see this, we use the fact that the weak topology on  $\mathcal{M}_+^f(E)$  can be metrised, for instance by the Kantorovich-Rubinstein metric,

$$d_{\text{KR}}(\mu, \nu) = \sup_{f \in \text{Lip}(1), \|f\|_\infty \leq 1} \int_E f d\mu - \int_E f d\nu.$$

It is easy to see that, for all bounded measurable  $f$ ,

$$\left| \int_E f d\Phi_n - \int_E f d\hat{\Phi}_n \right| \leq \|f\|_\infty |N_n - [nx]|,$$

and so,  $d_{\text{KR}}(\Phi_n/n, \hat{\Phi}_n/n)$  tends to zero deterministically, as  $N_n/n$  tends to  $x$  deterministically. This establishes the exponential equivalence of the two sequences.

Now, we have from Theorem 3.3 and the observation that  $\lambda_n(\cdot)/\lambda_n(E)$  converges weakly to  $\lambda(\cdot)/\lambda(E)$ , that  $(\hat{\Phi}_n/\lfloor nx \rfloor, \lfloor nx \rfloor \in \mathbb{N})$  obey an LDP in  $\mathcal{M}_1(E)$  with good rate function  $H(\cdot \mid \frac{\lambda}{\lambda(E)})$ , and hence also in  $\mathcal{M}_+^f(E)$  with rate function which is the same on  $\mathcal{M}_1(E)$ , and infinite outside it. It follows that  $(\hat{\Phi}_n/n, n \in \mathbb{N})$  obey an LDP in  $\mathcal{M}_+^f(E)$  with rate function

$$H_x(\mu) = \begin{cases} xH\left(\frac{\mu}{x} \mid \frac{\lambda}{\lambda(E)}\right), & \text{if } \frac{\mu}{x} \in \mathcal{M}_1(E), \\ +\infty, & \text{otherwise.} \end{cases} \quad (2)$$

Finally, by [5, Theorem 4.2.13],  $(\Phi_n/n, n \in \mathbb{N})$  obey an LDP in  $\mathcal{M}_+^f(E)$  with the same rate function  $H_x$ , as they are exponentially equivalent to  $\hat{\Phi}_n/n$ .

Having established conditional LDPs for  $\Phi_n/n$ , conditional on  $N_n/n$  tending to  $x$ , we now need to check the LDP continuity conditions in Definition 3.1 with  $\Omega_1 = \mathbb{R}_+$  and  $\Omega_2 = \mathcal{M}_+^f(E)$ , and transition function  $\nu_n(x, \cdot)$  defined as the law of  $\Phi_n$  conditional on  $N_n = \lfloor nx \rfloor$ . We define the function

$$J(x, \mu) = \begin{cases} \mathcal{I}_0(\mu), & \text{if } x = 0, \\ H_x(\mu) & \text{if } x > 0, \end{cases}$$

where  $\mathcal{I}_0$  is defined in Lemma 3.5 and  $H_x$  in (2). Note that  $J$  is non-negative as  $\mathcal{I}_0$  and  $\{H_x, x \geq 0\}$  are all non-negative.

The first condition in Definition 3.1 holds trivially if  $x = 0$ , as all level sets are singletons comprised of the zero measure; if  $x > 0$ , the condition follows from the goodness of the relative entropy function, which is well known from Sanov's theorem (see, e.g., [5, Theorem 6.2.10]). In a bit more detail, given  $\alpha > 0$ , the level set

$$L_\alpha = \left\{ \mu \in \mathcal{M}_1(E) : H\left(\mu \mid \frac{\lambda}{\lambda(E)}\right) \leq \frac{\alpha}{x} \right\}$$

is compact in  $\mathcal{M}_1(E)$  equipped with the weak topology; hence, so is its image under the continuous map  $\mu \mapsto x\mu$  from  $\mathcal{M}_1(E)$  to  $\mathcal{M}_+^f(E)$ .

The second condition in Definition 3.1 is precisely the content of the conditional LDPs that we just obtained. That leaves us to check the third condition, which is that  $J(x, \mu)$  is l.s.c. in  $(x, \mu)$ . As  $\mathbb{R}_+ \times \mathcal{M}_+^f(E)$  is a metric space, we can check this along sequences. Consider a sequence  $(x_n, \mu_n)$  converging to  $(x, \mu)$ . If  $(x, \mu) = (0, 0)$ , then  $J(x, \mu) = 0$ , which is no bigger than

$\liminf J(x_n, \mu_n)$ . If  $x = 0$  and  $\mu \neq 0$ , then  $\mu(E) > 0$  and so, for all  $n$  sufficiently large,  $x_n < \mu_n(E)$ ; consequently,  $\mu_n/x_n$  is not a probability measure, and  $J(x_n, \mu_n) = +\infty$ . The same reasoning applies if  $x > 0$  and  $\mu/x \notin \mathcal{M}_1(E)$ . Finally, suppose  $x > 0$  and  $\mu/x \in \mathcal{M}_1(E)$ , so that  $\mu_n/x_n$  converges weakly to  $\mu/x$  in  $\mathcal{M}_+^f(E)$ . We may restrict attention to the subsequence of  $\mathbb{N}$  for which  $\mu_n/x_n$  are probability measures, as  $J(x_n, \mu_n) = +\infty$  otherwise. Along this subsequence, the desired inequality  $\liminf H_{x_n}(\mu_n) \geq H_x(\mu)$  follows from the lower semicontinuity of  $H$ , the relative entropy function.

We are now in a position to invoke Theorem 3.2, with  $\Omega_1 = \mathbb{R}_+$  and  $\Omega_2 = \mathcal{M}_+^f(E)$ . The second condition in the theorem is a conditional LDP for  $\Phi_n/n$  given that  $N_n/n$  tends to  $x$ , which we have just verified. The first condition is an LDP for  $N_n/n$ , which was proved in Lemma 3.4. Hence, the conclusion of Theorem 3.2 holds, i.e., we have an LDP for  $\Phi_n/n$  with rate function

$$I_2(\mu) = \inf_{x \in \mathbb{R}_+} \{I_{Poi}(x, \lambda(E)) + J(x, \mu)\}.$$

As  $J(x, \mu) = +\infty$  unless  $x = \mu(E)$ , it is clear that the infimum is attained at  $x = \mu(E)$ , and we have

$$I_2(\mu) = I_{Poi}(\mu(E), \lambda(E)) + J(\mu(E), \mu).$$

This coincides with the rate function in the statement of the lemma, and concludes its proof.  $\square$

We now have all the ingredients required to complete the proof of Theorem 2.1.

**Proof of Theorem 2.1.** We invoke Theorem 3.2 with  $\Omega_1$  and  $\Omega_2$  both being the space of finite non-negative measures on  $E$ , equipped with the weak topology and the corresponding Borel  $\sigma$ -algebra. The sequence  $\mu_{1n}$  will denote the laws of the directing (intensity) measures  $\Lambda_n$ , and the probability transition functions  $\nu_n(\lambda, \cdot)$  will denote the law of the scaled Poisson random measures  $\Phi_n/n$ , where  $\Phi_n$  has intensity  $n\lambda$ . We now check the assumptions of the theorem.

The first condition in Theorem 3.2 is an LDP for  $(\Lambda_n/n, n \in \mathbb{N})$  with a good rate function, which holds by assumption. To check the second condition in Theorem 3.2, define

$$J(\lambda, \mu) = \begin{cases} \mathcal{I}_0(\mu), & \text{if } \lambda \equiv 0, \\ \mathcal{I}_1(\mu), & \text{otherwise,} \end{cases}$$

where  $\mathcal{I}_0$  and  $\mathcal{I}_1$  are as defined in Lemmas 3.5 and 3.6. We need to check that the conditions in Definition 3.1 are satisfied. The first condition is satisfied as

$\mathcal{I}_0$  and  $\mathcal{I}_1$  are both good rate functions, as shown in Lemmas 3.5 and 3.6. The second condition is the content of the conditional LDPs established in these lemmas. That leaves us to check the third condition, that  $J(\cdot, \cdot)$  is l.s.c.. As the weak topology on  $\mathcal{M}_+^f(E)$  is metrizable, so is the product topology on  $\mathcal{M}_+^f(E) \times \mathcal{M}_+^f(E)$ , and we can check lower semicontinuity along sequences. Consider a sequence  $(\lambda_n, \mu_n)$  converging to  $(\lambda, \mu)$ , i.e.,  $\lambda_n$  converges weakly to  $\lambda$ , and  $\mu_n$  to  $\mu$ . We distinguish four cases:

1. If  $\lambda \equiv 0$  and  $\mu \equiv 0$ , then  $J(\lambda, \mu) = \mathcal{I}_0(\mu) = 0$ , which is no bigger than the limit infimum of a non-negative sequence.
2. If  $\lambda \equiv 0$  and  $\mu \neq 0$ , then  $J(\lambda, \mu) = \mathcal{I}_0(\mu) = +\infty$ . But note that  $\lambda_n(E) \rightarrow \lambda(E) = 0$  and  $\mu_n(E) \rightarrow \mu(E) > 0$ , and so  $I_{Poi}(\mu_n(E), \lambda_n(E)) \rightarrow +\infty$ . As

$$J(\lambda_n, \mu_n) = \mathcal{I}_1(\mu_n) \geq I_{Poi}(\mu_n(E), \lambda_n(E)),$$

we see that  $J(\lambda_n, \mu_n)$  also tends to infinity.

3. If  $\lambda \neq 0$  and  $\mu \equiv 0$ , then  $J(\lambda, \mu) = \mathcal{I}_1(\mu) = I_{Poi}(0, \lambda(E))$ . On the other hand,  $J(\lambda_n, \mu_n) \geq I_{Poi}(\mu_n(E), \lambda_n(E))$ , which tends to  $I_{Poi}(0, \lambda(E))$  as  $n$  tends to infinity, as  $I_{Poi}$  is continuous.
4. Finally, suppose that  $\lambda \neq 0$  and  $\mu \neq 0$ . In this case, for all  $n$  sufficiently large, both  $\lambda_n$  and  $\mu_n$  are non-zero measures, and we have  $J(\lambda_n, \mu_n) = \mathcal{I}_1(\mu_n)$ . As  $\lambda_n(E)$  and  $\mu_n(E)$  converge to  $\lambda(E)$  and  $\mu(E)$  respectively, it is easy to see that  $I_{Poi}(\mu_n(E), \lambda_n(E))$  tends to  $I_{Poi}(\mu(E), \lambda(E))$ . Hence, to verify lower semicontinuity, it suffices to show that  $H(\beta|\alpha)$  is jointly l.s.c. in its arguments. Recall the Donsker-Varadhan variational formula for the relative entropy (see, e.g., [6, Sec. C.2]):

$$H(\beta|\alpha) = \sup_{g \in C_b(E)} \left\{ \int_E g d\beta - \log \int_E e^g d\alpha \right\},$$

where  $C_b(E)$  denotes the set of bounded continuous functions on  $E$ . But if  $g \in C_b(E)$ , so is  $e^g$ , and the map

$$(\alpha, \beta) \mapsto \int_E g d\beta - \log \int_E e^g d\alpha$$

is continuous. Consequently,  $H(\beta|\alpha)$ , being the supremum of continuous functions of  $(\alpha, \beta)$ , is l.s.c..



Thus, we have checked all the conditions of Theorem 3.2. Hence, the conclusion of the theorem holds, and yields that  $(\Phi_n/n, n \in \mathbb{N})$  obey an LDP on  $\mathcal{M}_+^f(E)$ , with rate function

$$\mathfrak{I}_2(\mu) = \inf_{\lambda \in \mathcal{M}_+^f(E)} \{ \mathfrak{I}_1(\lambda) + J(\lambda, \mu) \},$$

where  $J(\lambda, \mu)$  equals  $\mathcal{I}_0(\mu)$  if  $\lambda \equiv 0$  and  $\mathcal{I}_1(\mu)$  otherwise, and  $\mathcal{I}_0$  and  $\mathcal{I}_1$  are defined in Lemmas 3.5 and 3.6 respectively. Using those definitions, we can write the rate function more explicitly as follows:

$$\mathfrak{I}_2(\mu) = \begin{cases} \inf_{\lambda} \{ \mathfrak{I}_1(\lambda) + \lambda(E) \}, & \text{if } \mu \equiv 0, \\ \inf_{\lambda} \left\{ \mathfrak{I}_1(\lambda) + I_{Poi}(\mu(E), \lambda(E)) + \mu(E) H\left(\frac{\mu}{\mu(E)} \mid \frac{\lambda}{\lambda(E)}\right) \right\}, & \text{if } \mu \neq 0, \end{cases}$$

where the infimum is taken over all finite Borel measures  $\lambda$  on  $E$ . The expression above coincides with that in the statement of the theorem.

It remains only to check that the rate function  $\mathfrak{I}_2$  is good. This is a consequence of Lemma 3.8 below, which establishes the exponential tightness of the scaled empirical measures  $\Phi_n/n$ , and [5, Lemma 1.2.18]. This completes the proof of Theorem 2.1.  $\square$

We first state a proposition which provides an explicit construction of compact subsets of  $\mathcal{M}_+^f(E)$ , and which we will need for the proof of Lemma 3.8. The proof of the proposition is deferred until after the lemma, and is where the assumption of  $\sigma$ -compactness of  $E$  is required.

**Proposition 3.7.** *Let  $K_1 \subseteq K_2 \subseteq \dots$  be a nested sequence of compact subsets of  $E$ , whose union is equal to  $E$ ; such a sequence exists by the assumption that  $E$  is  $\sigma$ -compact. Let  $\varepsilon_0 \geq \varepsilon_1 \geq \dots$  be a sequence of real numbers decreasing to zero. Define  $K_0$  to be the empty set. Then, the set*

$$L_{(K_n, \varepsilon_n)} = \left\{ \mu \in \mathcal{M}_+^f(E) : \mu(K_n^c) \leq \varepsilon_n \ \forall \ n \in \mathbb{N} \right\},$$

*is compact in the weak topology on  $\mathcal{M}_+^f(E)$ . Moreover, if  $\mathcal{K}$  is any compact subset of  $\mathcal{M}_+^f(E)$ , and  $\varepsilon_n, n \in \mathbb{N}_+$  any sequence decreasing to 0, then there exist  $\varepsilon_0 > 0$  and compact  $K_1 \subseteq K_2 \subseteq \dots \subseteq E$  such that  $\mathcal{K} \subseteq L_{(K_n, \varepsilon_n)}$ .*

**Lemma 3.8.** *Suppose that  $(\Lambda_n, n \in \mathbb{N})$  is a sequence of random finite Borel measures on a Polish space  $(E, d)$ , which satisfy the assumptions of Theorem 2.1. Let  $(\Phi_n, n \in \mathbb{N})$  be a sequence of Cox point processes on  $E$ , with stochastic intensities  $\Lambda_n$ . Then, the sequence of random measures  $\Phi_n/n$  is exponentially tight in  $\mathcal{M}_+^f(E)$  equipped with the weak topology.*

*Proof.* We have to show that for every  $\alpha < \infty$ , there is a compact  $\mathcal{K}_\alpha \subseteq \mathcal{M}_+^f(E)$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n}{n} \in \mathcal{K}_\alpha^c\right) < -\alpha. \quad (3)$$

By the assumptions of Theorem 2.1, the sequence  $\Lambda_n/n$  satisfies an LDP in  $\mathcal{M}_+^f(E)$ , with *good* rate function  $\mathfrak{I}_1$ . Hence, the sequence is exponentially tight, i.e., there is a compact set  $\hat{\mathcal{K}}_\alpha \subseteq \mathcal{M}_+^f(E)$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Lambda_n}{n} \notin \hat{\mathcal{K}}_\alpha\right) < -\alpha. \quad (4)$$

By Proposition 3.7,  $\hat{\mathcal{K}}_\alpha$  is contained in a compact set of the form  $L_{(K_n, \varepsilon_n)}$ , where  $\varepsilon_n, n \geq 1$  can be chosen to decrease to zero arbitrarily. We will show that, for a suitably chosen sequence  $\delta_n \downarrow 0$ , the set  $L_{(K_n, \delta_n)}$  satisfies the upper bound in (3).

Observe that

$$\begin{aligned} \mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)}\right) &\leq \mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)} \mid \frac{\Lambda_n}{n} \in L_{(K_i, \varepsilon_i)}\right) + \mathbb{P}\left(\frac{\Lambda_n}{n} \notin L_{(K_i, \varepsilon_i)}\right) \\ &\leq \mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)} \mid \frac{\Lambda_n}{n} \in L_{(K_i, \varepsilon_i)}\right) + \mathbb{P}\left(\frac{\Lambda_n}{n} \notin \hat{\mathcal{K}}_\alpha\right). \end{aligned} \quad (5)$$

Now, conditional on  $\Lambda_n$ ,  $\Phi_n$  is a Poisson point process, and  $\Phi_n(K_i^c)$  is a Poisson random variable with mean  $\Lambda_n(K_i^c)$ . Thus, conditional on  $\Lambda_n/n \in L_{(K_i, \varepsilon_i)}$ , the random variable  $\Phi_n(K_i^c)$  is stochastically dominated by a Poisson random variable with mean  $n\varepsilon_i$ , for each  $i \in \mathbb{N}$ . Also, the event  $\{\Phi_n/n \notin L_{(K_i, \delta_i)}\}$  is the union of the events  $\{\Phi_n(K_i^c) > n\delta_i\}$  over  $i \in \mathbb{N}$ . Define  $m_n = \sup\{i : n\delta_i > 1\}$ . Since  $\Phi_n$  is a counting measure, the event  $\{\Phi_n(K_i^c) > n\delta_i\}$  coincides with  $\{\Phi_n(K_i^c) \geq 1\}$  for  $i > m_n$ . Hence, we obtain using the union bound that

$$\begin{aligned} \mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)} \mid \frac{\Lambda_n}{n} \in L_{(K_i, \varepsilon_i)}\right) &\leq \sum_{i=0}^{\infty} \mathbb{P}(Poi(n\varepsilon_i) > n\delta_i) \\ &= \sum_{i=0}^{m_n} \mathbb{P}(Poi(n\varepsilon_i) > n\delta_i) + \sum_{i=m_n+1}^{\infty} \mathbb{P}(Poi(n\varepsilon_i) \geq 1). \end{aligned} \quad (6)$$

Without loss of generality, we can take  $\varepsilon_0 \geq 1$ . Take  $\varepsilon_i = e^{-i}$  and  $\delta_i = \kappa/i$  for  $i \geq 1$ , for a constant  $\kappa$  to be determined, depending on  $\alpha$ . Take  $\delta_0 = \kappa\varepsilon_0$ . Then  $m_n = \lfloor \kappa n \rfloor$ , and we obtain using Markov's inequality that

$$\sum_{i=m_n+1}^{\infty} \mathbb{P}(Poi(n\varepsilon_i) \geq 1) \leq \sum_{i=\lfloor \kappa n \rfloor}^{\infty} ne^{-i} \leq \frac{ne^{-\kappa n}}{1 - e^{-1}}. \quad (7)$$

We also have the large deviations (Chernoff) bound for a Poisson random variable that, for  $\mu > \lambda$ ,

$$\mathbb{P}(\text{Poi}(\lambda) > \mu) \leq \exp\left(-\mu \log \frac{\mu}{\lambda} + \mu - \lambda\right),$$

from which it follows that

$$\mathbb{P}(\text{Poi}(n\varepsilon_i) > n\delta_i) \leq \begin{cases} \exp(-n\varepsilon_0(\kappa \log \kappa - \kappa + 1)), & i = 0, \\ \exp(-n\kappa \frac{\log \kappa + i - 1 - \log i}{i}), & i \geq 1. \end{cases}$$

Now,  $\varepsilon_0 \geq 1$  by assumption and, if  $\kappa$  is chosen sufficiently large, then it is easy to verify that  $(\log \kappa + i - 1 - \log i)/i$  is bigger than  $1/2$  for all  $i \geq 1$ . Hence, we obtain that

$$\sum_{i=0}^{m_n} \mathbb{P}(\text{Poi}(n\varepsilon_i) > n\delta_i) \leq e^{-n(\kappa \log \kappa - \kappa + 1)} + \kappa n e^{-\kappa n/2}, \quad (8)$$

as  $m_n = \lfloor \kappa n \rfloor$ . Substituting (7) and (8) in (6), we get

$$\mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)} \mid \frac{\Lambda_n}{n} \in L_{(K_i, \varepsilon_i)}\right) \leq \frac{ne^{-\kappa n}}{1 - e^{-1}} + e^{-n(\kappa \log \kappa - \kappa + 1)} + \kappa n e^{-\kappa n/2}.$$

It is clear from this that we can choose  $\kappa$  sufficiently large to ensure that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)} \mid \frac{\Lambda_n}{n} \in L_{(K_i, \varepsilon_i)}\right) \leq -\alpha. \quad (9)$$

Finally, combining (4), (5) and (9), we conclude that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Phi_n}{n} \notin L_{(K_i, \delta_i)}\right) \leq -\alpha.$$

This concludes the proof of the lemma.  $\square$

**Proof of Proposition 3.7.** The weak topology on the space of finite measures on a Polish space is metrizable [17]), and so it suffices to check sequential compactness. Let  $(\mu_n, n \in \mathbb{N})$  be a sequence of finite measures on  $E$  satisfying the assumptions of the proposition with respect to a nested sequence of compact sets  $K_n$  whose union is equal to  $E$ , and a sequence  $\varepsilon_n$  decreasing to zero. In particular, the measures are bounded;  $\mu_n(E) \leq \varepsilon_0$  for all  $n \in \mathbb{N}$ . We want to show that  $(\mu_n, n \in \mathbb{N})$  contains a convergent subsequence.

Recall that the space of subprobability measures on a compact set  $K$  is compact in the weak topology; this follows from the Banach-Alaoglu theorem

applied to the unit ball in the space of finite signed measures on  $K$ , which the Riesz representation theorem identifies with the dual of the Banach space  $C(K)$  of continuous functions on  $K$  equipped with the supremum norm. Hence, by Tychonoff's theorem, so is the space of finite measures on  $K$  bounded by an arbitrary constant  $\varepsilon_0$ .

Thus, the measures  $\mu_n$  restricted to  $K_1$  all lie within a compact set; hence, there is a subsequence  $\mu_{11}, \mu_{12}, \dots$ , whose restriction to  $K_1$  converges weakly to some  $\tilde{\mu}_1 \in \mathcal{M}_+^f(K_1)$ . Similarly, the restriction of this subsequence to  $K_2$  all lie within a compact set, and contain a convergent subsubsequence  $\mu_{21}, \mu_{22}, \dots$ . We can extend this reasoning to  $K_3, K_4$  and so on.

Formally, denote by  $p_n$  the projection from  $\mathcal{M}_+^f(E)$  to  $\mathcal{M}_+^f(K_n)$  and by  $p_{mn}$  the projection from  $\mathcal{M}_+^f(K_m)$  to  $\mathcal{M}_+^f(K_n)$  for  $m \geq n$ . Then, we can rewrite the above as:

$$p_1 \mu_{1n} \rightarrow \tilde{\mu}_1 \in \mathcal{M}_+^f(K_1), \quad p_2 \mu_{2n} \rightarrow \tilde{\mu}_2 \in \mathcal{M}_+^f(K_2), \quad \dots,$$

where the convergence is with respect to the weak topology on the corresponding spaces. Now consider the diagonal sequence  $\mu_{kk}$ . It is clear from the above that

$$p_n \mu_{kk} \xrightarrow{k \rightarrow \infty} \tilde{\mu}_n \in \mathcal{M}_+^f(K_n),$$

for each  $n$ . A natural question to ask is whether there is a measure  $\tilde{\mu} \in \mathcal{M}_+^f(E)$  such that  $\tilde{\mu}_n = p_n \tilde{\mu}$  for all  $n$ . The answer follows from a generalisation of Kolmogorov's Extension theorem by Yamasaki [18, Proposition 2.1]; it is affirmative if the measures  $\tilde{\mu}_n$  satisfy the consistency conditions  $p_{mn} \tilde{\mu}_m = \tilde{\mu}_n$  for all  $m > n$ . It is straightforward to verify these.

We now show that the diagonal subsequence  $\mu_{kk}$  converges weakly to the measure  $\tilde{\mu}$  (whose existence we have just shown) in the weak topology on  $\mathcal{M}_+^f(E)$ , and moreover that the limit  $\tilde{\mu}$  is in  $L_{(K_n, \varepsilon_n)}$ . We start with the latter. As  $\tilde{\mu}$  is a finite measure on the Polish space  $E$ , it is regular; therefore, as  $K_n$  are compact sets increasing to  $E$ ,  $\tilde{\mu}(K_n)$  increases to  $\tilde{\mu}(E)$ . Hence, for any  $m \in \mathbb{N}$ ,

$$\tilde{\mu}(K_m^c) = \lim_{n \rightarrow \infty} \tilde{\mu}(K_n) - \tilde{\mu}(K_m).$$

Now, for any fixed  $i > n > m$ ,  $\tilde{\mu}_i$  is the restriction (or projection) of  $\tilde{\mu}$  to the set  $K_i$ , and so

$$\tilde{\mu}(K_n) - \tilde{\mu}(K_m) = \tilde{\mu}_i(K_n) - \tilde{\mu}_i(K_m) \leq \tilde{\mu}_i(K_m^c) \leq \varepsilon_m.$$

The last inequality holds because  $\tilde{\mu}_i$  is the weak limit of measures whose mass on  $K_m^c$  is bounded by  $\varepsilon_m$ , and  $K_m^c$  is an open set. As this holds for

each  $n$ , we conclude on taking limits that  $\tilde{\mu}(K_m^c) \leq \varepsilon_m$ . But  $m$  was arbitrary, so  $\tilde{\mu} \in L_{(K_n, \varepsilon_n)}$ .

Next, given  $\delta > 0$  and a bounded continuous function  $g : E \rightarrow \mathbb{R}$ , choose  $\ell$  large enough that  $\varepsilon_\ell \|g\|_\infty < \delta$ . Next, pick  $m \geq \ell$  large enough that

$$\left| \int_{K_\ell} g d\mu_{\ell n} - \int_{K_\ell} g d\tilde{\mu}_\ell \right| \leq \delta \quad \forall n \geq m,$$

which is possible since  $\mu_{\ell n}$  converges weakly to  $\tilde{\mu}_\ell$  as  $n$  tends to infinity. Now,  $\mu_n$  is a subsequence of  $\mu_\ell$  for  $n \geq \ell$ , so the above inequality also holds for  $\int_{K_\ell} g(d\mu_{nn} - d\tilde{\mu}_\ell)$  for all  $n \geq m$ . Thus, we can write

$$\left| \int_E g d\mu_{nn} - \int_E g d\tilde{\mu} \right| \leq \left| \int_{K_\ell} g(d\mu_{nn} - d\tilde{\mu}_\ell) \right| + \left| \int_{K_\ell} g(d\tilde{\mu}_\ell - d\tilde{\mu}) \right| + 2\|g\|_\infty \varepsilon_\ell,$$

as  $\mu_{nn}(K_\ell^c)$  and  $\tilde{\mu}(K_\ell^c)$  are both bounded above by  $\varepsilon_\ell$ . We have just shown that the first integral above is smaller than  $\delta$  in absolute value, for all  $n \geq m$ . The second integral is zero as  $\tilde{\mu}_\ell$  is the restriction or projection of  $\tilde{\mu}$  to  $K_\ell$ . The last term is bounded by  $2\delta$  by the choice of  $\ell$ . Thus, we have shown that we can choose  $m$  in such a way that

$$\left| \int_E g d\mu_{nn} - \int_E g d\tilde{\mu} \right| \leq 3\delta$$

for all  $n \geq m$ . As  $g$  was an arbitrary bounded continuous function, this proves that  $\mu_{nn}$  converges to  $\tilde{\mu}$ . This completes the proof that  $L_{(K_n, \varepsilon_n)}$  is compact.

For the converse, let  $\mathcal{K}$  be compact in  $\mathcal{M}_+^f(E)$  equipped with the weak topology. As the map  $\mu \mapsto \mu(E)$  is continuous (the indicator of  $E$  is a bounded continuous function  $E \rightarrow \mathbb{R}$ ), its supremum over  $\mathcal{K}$  is attained. Denote the supremum by  $\varepsilon_0$ . Then  $\mu(E) = \mu(K_0^c) \leq \varepsilon_0$  for all  $\mu \in \mathcal{K}$ . Next, we invoke a generalisation of Prokhorov's theorem by Bogachev [3, Theorem 8.6.2]), which states that the measures in a compact set are uniformly tight. In other words, given  $\varepsilon_1 > 0$ , we can find a compact subset  $K_1$  of  $E$  such that  $\mu(K_1^c) \leq \varepsilon_1$  for all  $\mu \in \mathcal{K}$ . Similarly, we can find compact  $K_2$  such that  $\mu(K_2^c) \leq \varepsilon_2$  for all  $\mu \in \mathcal{K}$ . Without loss of generality, we can assume that  $K_1 \subseteq K_2$ ; otherwise, re-define  $K_2$  as their union. Continuing in the same vein, we obtain a sequence  $K_n$  of nested compact sets such that  $\mu(K_n^c) \leq \varepsilon_n$  for all  $n \in \mathbb{N}$ , for all  $\mu \in \mathcal{K}$ . If their union is not equal to  $E$ , it can be extended countably to have this property, by the assumption that  $E$  is  $\sigma$ -compact. Now,  $\mathcal{K} \subseteq L_{(K_n, \varepsilon_n)}$ .  $\square$

## 4 Proof of LDP for Queue Occupancy and Departures

The proof of Theorems 2.2 and 2.3 are presented in this section. We begin by recalling how the queue occupancy measure is related to the input to the queue. First, we represent the input to the  $n^{\text{th}}$  queue as a Cox process on  $\mathbb{R} \times \mathbb{R}_+$  by marking each arrival with its service time; the resulting marked point process is a Cox process on  $\mathbb{R} \times \mathbb{R}_+$  with stochastic intensity  $\Lambda_n \otimes F$ . Now,  $Q_n(t)$  is equal to the number of points of this Cox process lying in the triangle

$$A_t = \{(s, x) \in \mathbb{R} \times \mathbb{R}_+ : s \leq t, x \geq t - s\}.$$

Furthermore, the queue length process  $\{Q_n(t), t \in [a, b]\}$ , is determined by the restriction of the above Cox process to the wedge

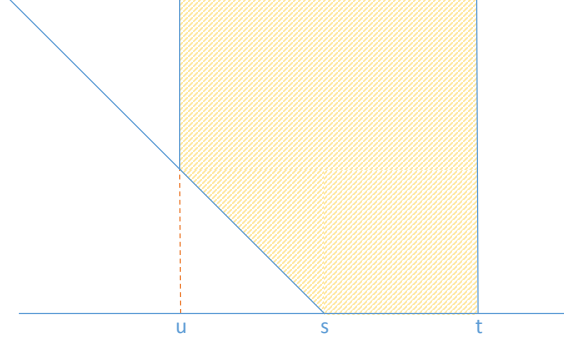
$$A_{[a,b]} := \bigcup_{t \in [a,b]} A_t,$$

as illustrated in Figure 1. Next, for  $u \leq s \leq t$ , we will also need to define the truncated sets

$$A_t^u = \{(s, x) \in \mathbb{R} \times \mathbb{R}_+ : u \leq s \leq t, x \geq t - s\}, \quad A_{[s,t]}^u := \bigcup_{x \in [s,t]} A_x^u.$$

Finally, recall that we are interested in the occupancy measure  $L_n$ , which is defined as the random measure that is absolutely continuous with respect to Lebesgue measure, and has density  $Q_n(\cdot)$ .

Our goal is to prove an LDP for  $L_n$ , restricted to an arbitrary interval  $[a, b]$ . We start by establishing an LDP for the scaled directing measures  $\frac{\Lambda_n}{n} \otimes F$ , restricted to a truncated wedge  $A_{[a,b]}^u$ , for arbitrary  $u < a$ ; we define below a new topology, which we call the tempered topology, in which we establish this LDP. Then, using the projective limit approach described below, we extend this family of LDPs to an LDP on the full wedge  $A_{[a,b]}$ , in the projective limit topology. However, the queueing map is not continuous in this topology, so we need to strengthen the LDP to the weak topology on the full wedge. We do this by establishing exponential tightness of the measures  $\frac{\Lambda_n}{n} \otimes F$  in the tempered topology on  $A_{[a,b]}$ . Next, we invoke Theorem 2.1 to deduce an LDP for the Cox process on  $A_{[a,b]}$  with this intensity. Finally, we use continuity of the queueing map with respect to the weak topology, and the contraction principle, to obtain the LDP for  $L_n$ . Checking that  $L_n$  also satisfies Assumptions [A1]-[A3] is fairly straightforward. The details of all these steps are presented below.



**Figure 1** – The wedge  $A_{[s,t]}$  and the truncated wedge  $A_{[s,t]}^u$ .

**Definition 4.1.** Let  $u \leq a < b \in \mathbb{R}$ , and let  $\mathcal{M}_+^f(A_{[a,b]}^u)$  denote the space of finite measures on the truncated wedge  $A_{[a,b]}^u$  defined above. The tempered topology on this space is the weakest topology which makes the maps  $\mu \mapsto \int f d\mu$  continuous for all bounded, continuous functions  $f : A_{[a,b]}^u \rightarrow \mathbb{R}$  which vanish at the boundary of  $A_{[a,b]}^u$ .

The tempered topology on  $\mathcal{M}_+^f(A_{[a,b]})$  is defined analogously.

Notice that the tempered topology is weaker than the weak topology, as it is restricted to test functions that vanish at the boundary. We are now ready to state our first result.

**Lemma 4.2.** Fix  $u \leq a < b \in \mathbb{R}$  and consider the truncated wedge  $A_{[a,b]}^u$ . The sequence of random measures  $\frac{\Delta_n}{n} \otimes F|_{A_{[a,b]}^u}$ ,  $n \in \mathbb{N}$ , satisfy an LDP on  $\mathcal{M}_+^f(A_{[a,b]}^u)$  equipped with the tempered topology, with good rate function

$$I_{[a,b]}^u(\mu) = \inf \left\{ I_{[a,b]}(\lambda) : \lambda \in \mathcal{M}_+^f([a,b]), \mu = (\lambda \otimes F)|_{A_{[a,b]}^u} \right\}.$$

*Proof.* Define the map

$$T : \mathcal{M}_+^f([u,b]) \rightarrow \mathcal{M}_+^f([u,b] \times \mathbb{R}_+)$$

by  $T(\mu) = \mu \otimes F$ . We first show that this map is continuous in the weak topology. As the weak topology is metrizable, we can check continuity along sequences. To this end, consider a sequence of finite measures  $\mu_n$  on  $[u,b]$  converging weakly to a finite measure  $\mu$ , and let  $g : [u,b] \times \mathbb{R}_+ \rightarrow \mathbb{R}$  be

bounded and continuous. Define  $h : [u, b] \mapsto \mathbb{R}$  by  $h(x) = \int_0^\infty g(x, y) dF(y)$ . We have

$$\int_{[u, b] \times \mathbb{R}_+} g d(T(\mu_n)) = \int_u^b \left( \int_0^\infty g(x, y) dF(y) \right) d\mu_n(x) = \int_u^b h(x) d\mu_n(x),$$

where the first equality follows from Fubini's theorem. If we can show that  $h$  is continuous, then it will follow that  $\int g d(T(\mu_n))$  converges to  $\int g d(T(\mu))$ , and, as  $g$  was an arbitrary bounded continuous function, that  $T(\mu_n)$  converges weakly to  $T(\mu)$ , thus proving that  $T$  is continuous.

Now, to show that  $h$  is continuous, fix  $\varepsilon > 0$  and  $x_0 \in \mathbb{R}$  such that  $1 - F(x_0) \leq \varepsilon$ . Now  $g$  is uniformly continuous on the compact set  $[u, b] \times [0, x_0]$ , so we can find  $\delta > 0$  such that  $|g(x, z) - g(y, z)| < \varepsilon$  provided  $|x - y| < \delta$ . It follows that

$$\begin{aligned} & |h(x) - h(y)| \\ & \leq \int_0^{x_0} |g(x, z) - g(y, z)| dF(z) + \int_{x_0}^\infty |g(x, z)| dF(z) + \int_{x_0}^\infty |g(y, z)| dF(z) \\ & \leq (1 + 2\|g\|_\infty)\varepsilon. \end{aligned}$$

This proves the continuity of  $h$ , and consequently of  $T$ .

Next, let  $S$  be the map that restricts finite measures on  $[u, b] \times \mathbb{R}_+$  to the wedge  $A_{[a, b]}^u$ . Equip  $\mathcal{M}_+^f([u, b] \times \mathbb{R}_+)$  with the weak topology, and  $\mathcal{M}_+^f(A_{[a, b]}^u)$  with the tempered topology. It is easy to see that  $S$  is continuous. Indeed, let  $\mu_n, n \in \mathbb{N}$  be a sequence of finite measures on  $[u, b] \times \mathbb{R}_+$  converging weakly to a finite measure  $\mu$  on  $[u, b] \times \mathbb{R}_+$ , and let  $f$  be a bounded, continuous function on  $A_{[a, b]}^u$ , vanishing on its boundary. Extend it to a bounded, continuous function  $\hat{f} : [u, b] \times \mathbb{R}_+ \rightarrow \mathbb{R}$  by defining  $\hat{f} \equiv f$  on  $A_{[a, b]}^u$  and  $\hat{f} \equiv 0$  on the complement of  $A_{[a, b]}^u$  in  $[u, b] \times \mathbb{R}_+$ . Then,

$$\int_{A_{[a, b]}^u} f d(S(\mu_n)) = \int_{[u, b] \times \mathbb{R}_+} \hat{f} d\mu_n \rightarrow \int_{[u, b] \times \mathbb{R}_+} \hat{f} d\mu = \int_{A_{[a, b]}^u} f d(S(\mu)),$$

where the convergence holds by the assumption that  $\mu_n$  converge weakly to  $\mu$ . This proves that  $S$  is continuous. As  $S$  and  $T$  are both continuous, so is the composition  $S \circ T$ . The claim of the lemma now follows from the assumed LDP for  $\frac{\Lambda_n}{n} \big|_{[u, b]}$  and the contraction principle [5, Theorem 4.2.1].  $\square$

The family of LDPs on the truncated wedges  $\{A_{[a, b]}^u, u < a\}$  can be extended to an LDP on the full wedge  $A_{[a, b]}$  using the Dawson-Gärtner theorem for projective limits [5, Theorem 4.6.1]. This yields an LDP in



the projective limit topology, which is generated by bounded continuous functions supported on the truncated wedges  $A_{[a,b]}^u$  and vanishing at their boundaries. In order to strengthen this LDP to the weak topology on  $A_{[a,b]}$ , we need to show exponential tightness of the measures  $\frac{\Lambda_n}{n} \otimes F$  in the weak topology. The following lemma is a key ingredient in establishing this.

**Lemma 4.3.** *Suppose  $X, X_1, X_2, \dots$  are identically distributed random variables with arbitrary joint distribution, and suppose  $\alpha_i, i \in \mathbb{N}$  are non-negative coefficients whose sum is finite, and which we denote by  $\alpha$ . Then,*

$$\sum_{i=1}^{\infty} \alpha_i X_i \leq_{\text{cx}} \alpha X,$$

where we write  $Y \leq_{\text{cx}} Z$  to denote that  $Y$  is dominated by  $Z$  in the convex stochastic order, i.e.,  $\mathbb{E}[\phi(Y)] \leq \mathbb{E}[\phi(Z)]$  for all convex functions  $\phi$  for which the expectations are defined, possibly infinite.

*Proof.* By scaling the random variables, we assume  $\alpha = 1$  without loss of generality. By Jensen's inequality, the inequality

$$\phi\left(\sum_{i=1}^{\infty} \alpha_i X_i(\omega)\right) \leq \sum_{i=1}^{\infty} \alpha_i \phi(X_i(\omega)),$$

holds pointwise on the probability space  $\Omega$ . Taking expectations on both sides yields the result if we can interchange expectation and summation on the right. We can certainly do so (by Tonelli's theorem) if the functions  $\phi$  are non-negative, and hence also if they are bounded below. Now, for any  $c \in \mathbb{R}$ , the function  $\phi_c$  defined by  $\phi_c(x) = \max\{c, \phi(x)\}$  is convex and bounded below, so we get

$$\mathbb{E}\left[\phi_c\left(\sum_{i=1}^{\infty} \alpha_i X_i\right)\right] \leq \sum_{i=1}^{\infty} \alpha_i \mathbb{E}[\phi_c(X_i)] = \left(\sum_{i=1}^{\infty} \alpha_i\right) \mathbb{E}[\phi_c(X)],$$

as the  $X_i$  are identically distributed with the same law as  $X$ . Since  $\phi \leq \phi_c$ , it follows that

$$\mathbb{E}\left[\phi\left(\sum_{i=1}^{\infty} \alpha_i X_i\right)\right] \leq \left(\sum_{i=1}^{\infty} \alpha_i\right) \mathbb{E}[\phi_c(X)],$$

for all  $c \in \mathbb{R}$ . Letting  $c$  decrease to  $-\infty$  on the right now yields the claim of the lemma. This can be justified by splitting  $\phi$  into its positive and negative parts, and using the monotone convergence theorem.  $\square$

We are now ready to show that the directing measures restricted to a wedge are exponentially tight in the weak topology.

**Proposition 4.4.** *The sequence of random measures*

$$\left( \left( \frac{\Lambda_n}{n} \otimes F \right) \Big|_{A_{[a,b]}} \right)_{n \in \mathbb{N}}$$

*is exponentially tight in the weak topology.*

*Proof.* We have to show that for every  $0 < \alpha < \infty$ , there is a compact set  $\mathcal{K}_\alpha \subseteq \mathcal{M}_+^f(A_{[a,b]})$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \left( \frac{\Lambda_n}{n} \otimes F \right) \Big|_{A_{[a,b]}} \in \mathcal{K}_\alpha^c \right) < -\alpha. \quad (10)$$

We will use the explicit construction of a weakly compact set of measures given in Proposition 3.7. We seek a nested sequence of compact sets  $K_1 \subseteq K_2 \subseteq \dots \subseteq A_{[a,b]}$ , whose union is the wedge  $A_{[a,b]}$ , and a sequence of positive constants  $\varepsilon_0 \geq \varepsilon_1 \geq \dots$  decreasing to zero, such that

$$\mathbb{P} \left( \left( \frac{\Lambda_n}{n} \otimes F \right) (K_i^c) > \varepsilon_i \right) \leq e^{-n(i+1)\alpha} \quad \forall i \geq 0, \quad (11)$$

where we define  $K_0$  to be the empty set. If we can find such  $K_i$  and  $\varepsilon_i$ , then the weakly compact set of measures

$$\mathcal{K}_\alpha = \left\{ \mu \in \mathcal{M}_+^f(A_{[a,b]}) : \mu(K_i^c) \leq \varepsilon_i \quad \forall i \in \mathbb{N} \right\},$$

satisfies the inequality in (10), thus proving the proposition.

Each of the compact sets  $K_i$ ,  $i \geq 1$ , will be specified by two real numbers  $u_i$  and  $h_i$  as shown in Figure 2:

$$K_i = \{[u_i, b] \times [0, h_i]\} \cap A_{[a,b]}.$$

We shall write  $K_i^c$  to denote the complement of  $K_i$  in  $A_{[a,b]}$ , and we decompose this set into a triangle

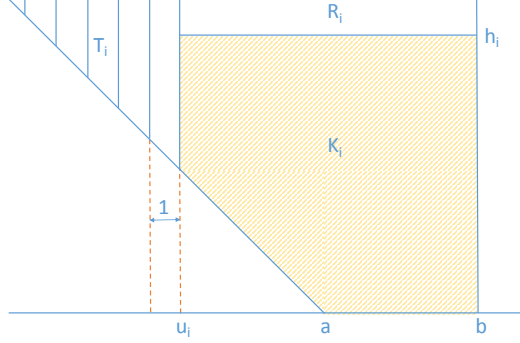
$$T_i = \{(s, x) \in \mathbb{R} \times \mathbb{R}_+ : s \leq u_i, x \geq a - s\},$$

and a rectangle

$$R_i = \{(s, x) \in \mathbb{R} \times \mathbb{R}_+ : u_i \leq s \leq b, x \geq h_i\};$$

see Figure 2. Thus, we have

$$\frac{1}{n} (\Lambda_n \otimes F)(K_i^c) = \frac{1}{n} (\Lambda_n \otimes F)(T_i) + \frac{1}{n} (\Lambda_n \otimes F)(R_i). \quad (12)$$



**Figure 2** – The wedge  $A_{[a,b]}$  split into a compact set  $K_i$ , infinite rectangle  $R_i$  and infinite triangle  $T_i$ . The triangle is split into strips of unit width.

Now, by the translation invariance of  $\Lambda_n$ , we have

$$(\Lambda_n \otimes F)(T_i) \stackrel{d}{=} (\Lambda_n \otimes F)(T^{a-u_i}) \text{ and } (\Lambda_n \otimes F)(R_i) \stackrel{d}{=} (\Lambda_n \otimes F)(R_{b-u_i}^{h_i}),$$

where  $\stackrel{d}{=}$  denotes equality in distribution, and the sets  $T^\ell$  and  $R_z^h$  are defined as

$$\begin{aligned} T^\ell &= \{(t, x) \in \mathbb{R} \times \mathbb{R}_+ : t \leq 0, t + x \geq \ell\} \\ R_z^h &= \{(t, x) \in \mathbb{R} \times \mathbb{R}_+ : t \in [0, z], x \geq h\}. \end{aligned} \quad (13)$$

Thus, we obtain from (12) that

$$\begin{aligned} \mathbb{P}\left(\left(\frac{\Lambda_n}{n} \otimes F\right)(K_i^c) > \varepsilon_i\right) &\leq \mathbb{P}\left((\Lambda_n \otimes F)(T^{a-u_i}) > \frac{n\varepsilon_i}{2}\right) \\ &\quad + \mathbb{P}\left((\Lambda_n \otimes F)(R_{b-u_i}^{h_i}) > \frac{n\varepsilon_i}{2}\right). \end{aligned} \quad (14)$$

We show in Lemma 4.5 that, given  $i \in \mathbb{N}$ ,  $\varepsilon_i > 0$  and  $\alpha > 0$ , we can choose  $u_i$  to make  $a - u_i$  sufficiently large that

$$\mathbb{P}\left((\Lambda_n \otimes F)(T^{a-u_i}) > \frac{n\varepsilon_i}{2}\right) \leq e^{-n(i+1)\alpha}, \quad \forall n \in \mathbb{N};$$

to see this, take  $\varepsilon = \varepsilon_i/2$  and  $\beta = (i+1)\alpha$  in the statement of the lemma. Next, by the same lemma, given  $u_i$ , and hence  $b - u_i$ , we can choose  $h_i$  sufficiently large to ensure that

$$\mathbb{P}\left((\Lambda_n \otimes F)(R_{b-u_i}^{h_i}) > \frac{n\varepsilon_i}{2}\right) \leq e^{-n(i+1)\alpha}, \quad \forall n \in \mathbb{N}.$$

Combining these two inequalities with (14), we conclude that for all  $i \geq 1$ ,

$$\mathbb{P}\left((\Lambda_n \otimes F)(K_i^c) > n\varepsilon_i\right) \leq 2e^{-n(i+1)\alpha}, \quad \forall n \in \mathbb{N}, \quad (15)$$

which is essentially the same as (11). That leaves the case  $i = 0$ .

The same argument does not work for  $K_0$  as we cannot choose this set;  $K_0$  is the empty set and  $K_0^c = A_{[a,b]}$ . Instead, we need to show that we can choose  $\varepsilon_0$  sufficiently large that

$$\mathbb{P}\left((\Lambda_n \otimes F)(A_{[a,b]}) > n\varepsilon_0\right) \leq e^{-n\alpha}, \quad \forall n \in \mathbb{N}. \quad (16)$$

We first note that  $A_{[a,b]} \subset T_0 \cup \{[a-\ell, b] \times \mathbb{R}_+\}$ , where

$$T_0 = \{(t, x) \in \mathbb{R} \times \mathbb{R}_+ : t \leq a - \ell, t + x \geq a\}.$$

Hence

$$(\Lambda_n \otimes F)(A_{[a,b]}) \leq (\Lambda_n \otimes F)(T_0) + \Lambda_n([a - \ell, b]).$$

Moreover, by translation invariance of  $\Lambda_n$ , we have

$$(\Lambda_n \otimes F)(T_0) \stackrel{d}{=} (\Lambda_n \otimes F)(T^\ell),$$

where  $T^\ell$  is defined in (13). Using Lemma 4.5 below, we conclude that we can choose  $\ell$  sufficiently large that

$$\mathbb{P}\left((\Lambda_n \otimes F)(T_0) > n\right) \leq e^{-n\alpha}, \quad \forall n \in \mathbb{N}. \quad (17)$$

We also see from the proof of Lemma 4.5 that  $\Lambda_n([a - \ell, b])$  is dominated, in the increasing convex order, by  $\lceil \ell + b - a \rceil \Lambda_n([0, 1])$ ; in particular,

$$\mathbb{E}\left[e^{\theta \Lambda_n([a-\ell, b])}\right] \leq \mathbb{E}\left[e^{\theta(\ell+1+b-a)\Lambda_n([0,1])}\right] = \exp\left(\psi_n(n\theta(\ell+1+b-a))\right),$$

where  $\psi_n$  is defined in Assumption [A3]. By [A3], for given  $a, b, \ell$ ,  $\psi_n(n\theta(\ell+1+b-a))/n$  is bounded, for  $\theta$  in a neighbourhood of the origin, uniformly in  $n$ , i.e., there exist constants  $\theta, \delta > 0$  such that  $\psi_n(n\theta) \leq n\delta$  for all  $n \in \mathbb{N}$ . Consequently, by Markov's inequality,

$$\mathbb{P}(\Lambda_n([a - \ell, b]) \geq n(\varepsilon_0 - 1)) \leq e^{-n\theta(\varepsilon_0 - 1) + n\delta}, \quad \forall n \in \mathbb{N}.$$

Clearly, we can choose  $\varepsilon_0$  large enough to ensure that

$$\mathbb{P}(\Lambda_n([a - \ell, b]) \geq n(\varepsilon_0 - 1)) \leq e^{-n\alpha}, \quad \forall n \in \mathbb{N}.$$

Combining the above equation with (17), we see that the inequality in (16) holds, up to a factor of two. This completes the proof that the inequality in (11) holds for all  $i \geq 0$ , up to a factor of two on the RHS. Now, using the union bound over  $i$ , we get

$$\mathbb{P}\left(\exists i \geq 0 : \left(\frac{\Lambda_n}{n} \otimes F\right)(K_i^c) > \varepsilon_i\right) \leq \sum_{i=0}^{\infty} e^{-n(i+1)\alpha} \leq 2e^{-n\alpha},$$

from which (10) is immediate, given the definition of  $\mathcal{K}_\alpha$ . This completes the proof of the proposition.  $\square$

**Lemma 4.5.** *Let  $\beta > 0$  be a given constant. For  $\ell, h, z > 0$ , let the triangle  $T_\ell$  and the rectangle  $R_z^h$  be defined as in (13). Then, we have the following:*

1. *Given  $\varepsilon > 0$ , we can choose  $\ell$  sufficiently large that*

$$\mathbb{P}\left((\Lambda_n \otimes F)(T^\ell) > n\varepsilon\right) \leq e^{-n\beta}, \quad \forall n \in \mathbb{N}.$$

2. *Given  $z > 0$  and  $\varepsilon > 0$ , we can choose  $h$  sufficiently large that*

$$\mathbb{P}\left((\Lambda_n \otimes F)(R_z^h) > n\varepsilon\right) \leq e^{-n\beta}, \quad \forall n \in \mathbb{N}.$$

*Proof.* Fix an  $\ell \in \mathbb{R}$ . By splitting the triangle  $T^\ell$  into vertical strips of unit width, we see that

$$(\Lambda_n \otimes F)(T^\ell) \leq \sum_{k=0}^{\infty} \Lambda_n([-k-1, -k]) \bar{F}(\ell + k).$$

Now, by translation invariance of  $\Lambda_n$ , the random variables  $\Lambda_n([-k-1, -k])$  are identically distributed for all  $k$ . Moreover, the sum of the coefficients  $\bar{F}(\ell + k)$  can be bounded as follows:

$$\sum_{k=0}^{\infty} \bar{F}(\ell + k) \leq c_\ell := \int_{\ell-1}^{\infty} \bar{F}(x) dx = \mathbb{E}[S \mathbf{1}(S \geq \ell - 1)],$$

where  $S$  denotes a random variable with the distribution  $F$  of the service time, and  $\mathbf{1}(E)$  denotes the indicator of the event  $E$ . This last expectation is finite by the assumption that the service time has finite mean. Hence, invoking Lemma 4.3, we obtain that

$$(\Lambda_n \otimes F)(T^\ell) \leq_{icx} c_\ell \Lambda_n([0, 1]),$$

where, for random variables  $X$  and  $Y$ , we say that  $X$  is dominated by  $Y$  in the increasing convex order, written  $X \leq_{icx} Y$ , if  $\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)]$  for all increasing convex functions  $\phi$ . Applying this bound to the increasing convex function  $\phi(x) = e^{\theta x}$  for arbitrary  $\theta > 0$ , and using Markov's inequality, we get, for any  $\varepsilon > 0$ ,

$$\mathbb{P}\left((\Lambda_n \otimes F)(T^\ell) \geq \frac{n\varepsilon}{2}\right) \leq e^{-n\theta\varepsilon/2} \mathbb{E}\left[e^{\theta c_\ell \Lambda_n([0,1])}\right] = \exp\left(-\frac{n\theta\varepsilon}{2} + \psi_n(n\theta c_\ell)\right),$$

where the function  $\psi_n$  was defined in Assumption [A3]. As  $\theta > 0$  is arbitrary, it is convenient to rewrite the above inequality, replacing  $\theta$  by  $\theta/c_\ell$ , as

$$\log \mathbb{P}\left((\Lambda_n \otimes F)(T^\ell) \geq \frac{n\varepsilon}{2}\right) \leq -\frac{n\theta\varepsilon}{2c_\ell} + \psi_n(n\theta), \text{ where } c_\ell = \mathbb{E}[S\mathbf{1}(S \geq \ell - 1)]. \quad (18)$$

Now, by Assumption [A3], there exist positive constants  $\delta$  and  $\theta$  such that  $\psi_n(n\theta) \leq n\delta$ , uniformly in  $n$ . Moreover, as  $\mathbb{E}[S]$  is finite by Assumption [A4], it follows that  $c_\ell$  tends to zero as  $\ell$  tends to infinity. Hence, we see from (18) that, given  $i \in \mathbb{N}$  and  $\beta, \varepsilon > 0$ , we can choose  $\ell$  sufficiently large, and consequently  $c_\ell$  sufficiently small, to ensure that

$$\mathbb{P}\left((\Lambda_n \otimes F)(T^\ell) \geq n\varepsilon\right) \leq e^{-n\beta} \quad \forall n \in \mathbb{N}. \quad (19)$$

This completes the proof of the first claim of the lemma.

The proof of the second claim is very similar. We show that

$$(\Lambda_n \otimes F)(R_{b-a}^h) \leq_{icx} [b-a]\overline{F}(h)\Lambda_n([0,1]),$$

and apply Markov's inequality to the exponential of the random variable on the RHS. The details are omitted.  $\square$

We now have all the ingredients required to establish an LDP for the scaled intensity measures  $(\Lambda_n \otimes F)/n$ , on the wedge  $A_{[a,b]}$ .

**Proposition 4.6.** *Suppose that  $\Lambda_n, n \in \mathbb{N}$  is a sequence of random measures satisfying Assumptions [A1]-[A3] and  $F$  satisfies [A4]. Fix an interval  $[a, b] \subset \mathbb{R}$ . The sequence of random measures  $(\frac{\Lambda_n}{n} \otimes F)|_{A_{[a,b]}}$ ,  $n \in \mathbb{N}$ , satisfy an LDP on  $\mathcal{M}_+^f(A_{[a,b]})$  equipped with the weak topology, with good rate function*

$$I_{[a,b]}(\nu) = \sup_{u \leq a} I_{[a,b]}^u(\nu|_{A_{[a,b]}^u}), \quad \nu \in \mathcal{M}_+^f([a, b]).$$

*Proof.* We will use the Dawson-Gärtner theorem [5, Theorem 4.6.1] for projective limits. Letting

$$J := \left\{ A_{[a,b]}^u : u \in (-\infty, a) \right\},$$

it is clear that the collection  $(J, \subseteq)$  of truncated wedges  $A_{[a,b]}^u$  equipped with set inclusion is totally ordered, and hence also right-filtering. The set is indexed by  $u$ , and we will use  $u$  to denote the element  $A_{[a,b]}^u$ , to simplify notation. Denote by  $\mathcal{Y}_u$  the space  $\mathcal{M}_+^f(A_{[a,b]}^u)$  of finite measures on  $A_{[a,b]}^u$ , equipped with the tempered topology.

If  $t \leq u$ , i.e.,  $A_{[a,b]}^u \subseteq A_{[a,b]}^t$  (note that the order in the projective system reverses inequalities from the order on the real line), define the projection  $p_{ut} : \mathcal{Y}_t \rightarrow \mathcal{Y}_u$  by the restriction of a measure on  $A_{[a,b]}^t$  to the subset  $A_{[a,b]}^u$ . It is clear that this map is continuous in the tempered topology, since any bounded, continuous function on  $A_{[a,b]}^u$ , vanishing on its boundary, can be extended to a bounded, continuous function on  $A_{[a,b]}^t$ , vanishing on its boundary, by setting it to zero outside  $A_{[a,b]}^u$ . Moreover, the projections satisfy the consistency condition  $p_{us} = p_{ut} \circ p_{ts}$  for  $s \leq t \leq u$ . Thus,  $(\mathcal{Y}_u, p_{ut})_{t \leq u}$  constitute a projective system. We can identify  $\mathcal{M}_+^f(A_{[a,b]})$  with the projective limit, with canonical projections

$$p_u : \mathcal{M}_+^f(A_{[a,b]}) \rightarrow \mathcal{M}_+^f(A_{[a,b]}^u)$$

defined as the restriction of a measure from the full wedge  $A_{[a,b]}$  to its truncation  $A_{[a,b]}^u$ . These are clearly continuous in the tempered topology, by the same argument as above.

Now, by Lemma 4.2, the projections

$$\left( \frac{\Lambda_n}{n} \otimes F \right) \Big|_{A_{[a,b]}^u} = p_u \left( \left( \frac{\Lambda_n}{n} \otimes F \right) \Big|_{A_{[a,b]}} \right), \quad n \in \mathbb{N},$$

satisfy an LDP for each  $u \in (-\infty, a)$ , with rate function  $I_{[a,b]}^u$ . Hence, by the Dawson-Gärtner theorem, the sequence of measures  $\left( \frac{\Lambda_n}{n} \otimes F \right) \Big|_{A_{[a,b]}}$ ,  $n \in \mathbb{N}$ , satisfies an LDP in the projective limit topology, with good rate function

$$I_{[a,b]}(\nu) = \sup_{u \leq a} I_{[a,b]}^u(\nu \Big|_{A_{[a,b]}^u}), \quad \nu \in \mathcal{M}_+^f([a, b]).$$

Moreover, by Proposition 4.4, the measures  $\left( \frac{\Lambda_n}{n} \otimes F \right) \Big|_{A_{[a,b]}}$  are exponentially tight in the weak topology on  $\mathcal{M}_+^f(A_{[a,b]})$ . Hence, by [5, Corollary 4.2.6], we obtain that the LDP holds in the weak topology. Exponential tightness also implies goodness of the rate function [5, Lemma 1.2.18].  $\square$

Next, we show the continuity of the queueing map, which is the prelude to obtaining the LDP for the queue occupancy measure. For a measure  $\nu \in \mathcal{M}_+^f(A_{[a,b]})$ , and  $t \in [a, b]$ , we define  $Q^\nu(t) = \nu(A_t)$ , where we recall that  $A_t = A_{[t,t]}$  is the set

$$\{(s, x) \in \mathbb{R} \times \mathbb{R}_+ : s \leq t, s + x \geq t\}.$$

The interpretation is that, if  $\nu$  is a counting measure representing the marked arrival process into an infinite-server queue, where each arrival is marked with its service time, then  $Q^\nu(t)$  denotes the number of customers in the queue at time  $t$ . Let  $L(\nu)$  denote the measure on  $[a, b]$  which is absolutely continuous with respect to Lebesgue measure, and has density  $Q^\nu(\cdot)$ ; let  $L$  denote the map from  $\mathcal{M}_+^f(A_{[a,b]})$  to  $\mathcal{M}_+^f([a, b])$  which takes  $\nu$  to  $L(\nu)$ .

We want an explicit characterisation of the map  $L$ . We will describe  $L(\nu)$  through its action on the dual space  $C_b([a, b])$  of bounded, continuous functions on  $[a, b]$ , i.e., by specifying  $\int_a^b g(t) dL(\nu)(t)$  for all  $g \in C_b([a, b])$ . By the Riesz representation theorem,  $L(\nu)$  is uniquely determined by these integrals. From the description above, we have

$$\begin{aligned} \int_a^b g(t) dL(\nu)(t) &= \int_a^b g(t) Q^\nu(t) dt = \int_{t=a}^b g(t) \nu(A_t) dt \\ &= \int_{A_{[a,b]}} \left( \int_{\max\{a,s\}}^{\min\{s+x,b\}} g(t) dt \right) \nu(ds \times dx). \end{aligned} \quad (20)$$

The last equality is obtained by interchanging the order of integration, noting that an area element at  $ds \times dx$  contributes to  $\nu(A_t)$  for each  $t$  between  $\max\{a, s\}$  and  $\min\{s + x, b\}$ .

**Lemma 4.7.** *The map  $L : \mathcal{M}_+^f(A_{[a,b]}) \rightarrow \mathcal{M}_+^f([a, b])$ , defined by (20) via the Riesz representation theorem, is continuous with respect to the weak topology on each of these sets.*

*Proof.* The weak topology on the space of finite measures on a Polish space is metrizable [17], so we can check continuity of  $L$  along sequences. Suppose  $\nu_n, n \in \mathbb{N}$  converge to  $\nu$  in the weak topology on  $\mathcal{M}_+^f(A_{[a,b]})$ . Let  $g : [a, b] \rightarrow \mathbb{R}$  be a bounded, continuous function. We have by (20) that

$$\begin{aligned} \int_a^b g(t) dL(\nu_n)(t) &= \int_{A_{[a,b]}} h(s, x) \nu_n(ds \times dx), \\ \text{where} \quad h(s, x) &= \int_{\max\{a,s\}}^{\min\{s+x,b\}} g(t) dt, \end{aligned} \quad (21)$$



where the last integral is defined to be zero if the upper limit of integration is smaller than the lower limit. (In other words, the domain of integration should be understood to be empty in this case, rather than treating it as a signed integral with limits reversed.)

It is clear that the function  $h : A_{[a,b]} \rightarrow \mathbb{R}$  is bounded and continuous. Hence, it follows from the assumed convergence of  $\nu_n$  to  $\nu$  in the weak topology that the RHS in (21) converges to

$$\int_{A_{[a,b]}} h(s, x) \nu(ds \times dx).$$

This completes the proof of the lemma.  $\square$

We are now ready to prove the main result.

**Proof of Theorem 2.2.** Let  $\Phi_n$  denote the Cox process of arrivals into the  $n^{\text{th}}$  queue, marked with their service times. Fix  $[a, b] \subset \mathbb{R}$ . By Proposition 4.6, the sequence of measures  $(\frac{\Lambda_n}{n} \otimes F)|_{A_{[a,b]}}$ , satisfy an LDP on  $\mathcal{M}_+^f(A_{[a,b]})$  equipped with the weak topology, with good rate function  $I_{[a,b]}$  given therein. Hence, by Theorem 2.1, the sequence of Cox point measures  $\frac{\Phi_n}{n}|_{A_{[a,b]}}$  also satisfies an LDP on  $\mathcal{M}_+^f(A_{[a,b]})$  equipped with the weak topology, with good rate function  $\mathcal{I}_{[a,b]}$  given by

$$\mathcal{I}_{[a,b]}(\mathbf{0}) = \inf_{\lambda} \{I_{[a,b]}(\lambda) + \lambda(A_{[a,b]})\}, \quad (22)$$

where  $\mathbf{0}$  denotes the zero measure, whereas, for  $\mu \neq \mathbf{0}$ ,

$$\begin{aligned} \mathcal{I}_{[a,b]}(\mu) &= \inf_{\lambda} \left\{ I_{[a,b]}(\lambda) + I_{Poi}(\mu(A_{[a,b]}), \lambda(A_{[a,b]})) \right. \\ &\quad \left. + \mu(A_{[a,b]}) H\left(\frac{\mu}{\mu(A_{[a,b]})} \mid \frac{\lambda}{\lambda(A_{[a,b]})}\right) \right\}, \end{aligned} \quad (23)$$

where  $H$  and  $I_{Poi}$  are defined in the statements of Theorem 3.3 and Lemma 3.4 respectively.

Now, the queue occupancy measures  $L_n$  are given by  $L_n/n = L(\Phi_n/n)$ , where the map  $L$  is defined by (20), and is linear and weakly continuous. Hence, by the contraction principle [5, Theorem 4.2.1], the sequence of measures  $L_n/n$  satisfies an LDP on  $\mathcal{M}_+^f([a, b])$  equipped with the weak topology, with good rate function

$$J_{[a,b]}(\nu) = \inf \{ \mathcal{I}_{[a,b]}(\mu) : L(\mu) = \nu \}, \quad (24)$$

where the infimum of an empty set is defined to be  $+\infty$ . Thus, the sequence  $L_n$  satisfies Assumption [A2]. The measures  $L_n$  inherit translation

invariance from  $\Lambda_n$  via  $\Lambda_n \otimes F$  and  $\Phi_n$ , while finiteness of the mean follows easily from that of  $\lambda$  (the mean arrival intensity) and of the service time distribution. Thus, [A1] is verified. It remains to check [A3].

Observe that, analogous to (21), we have

$$\begin{aligned} L_n([0, 1]) &= (L(\Phi_n))([0, 1]) \\ &= \int_{(s, x) \in A_{[0, 1]}} (\min\{s + x, 1\} - \max\{s, 0\}) \Phi_n(ds \times dx) \\ &\leq \Phi_n(A_{[0, 1]}). \end{aligned}$$

But, conditional on  $\Lambda_n \equiv \lambda$ ,  $\Phi_n([0, 1])$  is a Poisson random variable with mean  $(\lambda \otimes F)(A_{[0, 1]})$ . Hence, we have for  $\theta \geq 0$  that

$$\mathbb{E} \left[ e^{\theta L_n([0, 1])} \right] \leq \mathbb{E} \left[ \exp \left( (e^\theta - 1) (\Lambda_n \otimes F)(A_{[0, 1]}) \right) \right].$$

Moreover, it can be shown by splitting  $A_{[0, 1]}$  into vertical strips of unit width and invoking Lemma 4.3, as in the proof of Lemma 4.5, that

$$(\Lambda_n \otimes F)(A_{[0, 1]}) \leq_{icx} (1 + \mathbb{E}[S]) \Lambda_n([0, 1]),$$

where  $\mathbb{E}[S]$  denotes the mean service time, and is finite by Assumption [A4]. Hence, we obtain for  $\theta \geq 0$  that

$$\mathbb{E} \left[ e^{\theta L_n([0, 1])} \right] \leq \mathbb{E} \left[ \exp \left( (e^\theta - 1) (1 + \mathbb{E}[S]) (\Lambda_n([0, 1])) \right) \right].$$

By Assumption [A3], there is a neighbourhood of 0 on which

$$\frac{\psi_n(n\eta)}{n} = \frac{1}{n} \log \mathbb{E} \left[ e^{\eta \Lambda_n([0, 1])} \right]$$

is bounded, uniformly in  $n$ . Setting  $\eta = (e^\theta - 1)(1 + \mathbb{E}[S])$ , we obtain uniform boundedness of

$$\frac{1}{n} \log \mathbb{E} \left[ e^{\theta L_n([0, 1])} \right]$$

for  $\theta \geq 0$  sufficiently small, uniformly in  $n$ . Boundedness is automatic for  $\theta < 0$  as the random variables  $L_n([0, 1])$  are non-negative. Thus, the sequence of measures  $L_n$  satisfy [A3] as well. This completes the proof of the theorem.  $\square$

Having established the LDP for the queue occupancy measure, we now turn to the empirical measure of the departure process from the infinite-server queue, which can be expressed as a function of the marked arrival

process, where the marks specify the service times. Fix a compact interval  $[a, b] \subset \mathbb{R}$ , and let  $D$  denote the function which maps the marked arrival process on  $A_{[a,b]}$ , to the departure process measure on  $[a, b]$ , as described in (1). We will formally define  $D$  via the Riesz representation theorem, by specifying, for each  $\nu \in \mathcal{M}_+^f(A_{[a,b]})$ , the integral with respect to  $D(\nu)$  of arbitrary bounded, continuous functions on  $[a, b]$ . Let  $g \in C_b([a, b])$  be one such function. We define the function  $h_g$  on  $A_{[a,b]}$  by setting

$$h_g(s, x) = \begin{cases} g(s + x), & (s, x) \in \text{cl}(A_{[a,b]} \setminus A_b), \\ 0, & \text{otherwise,} \end{cases} \quad (25)$$

and define the map  $\nu \mapsto D(\nu)$  by setting

$$\int_{[a,b]} g d(D(\nu)) = \int_{A_{[a,b]}} h_g d\nu, \quad \forall g \in C_b([a, b]). \quad (26)$$

It is clear from (1) that

$$\int_{[a,b]} g d\Psi_n = \int_{A_{[a,b]}} h_g d\hat{\Phi}_n, \quad \forall g \in C_b([a, b]),$$

i.e.,  $\Psi_n = D(\hat{\Phi}_n)$ . We will show that  $D(\cdot)$  is continuous in a suitable topology, and use this to establish the desired LDP for  $(\Psi_n, n \in \mathbb{N})$ .

**Proof of Theorem 2.3.** We begin by showing that the map  $D : \mathcal{M}_+^f(A_{[a,b]}) \rightarrow \mathcal{M}_+^f([a, b])$  defined by (26) is continuous, when  $\mathcal{M}_+^f(A_{[a,b]})$  is equipped with the weak topology, and  $\mathcal{M}_+^f([a, b])$  with the tempered topology. We can check continuity using sequences, as the weak topology on  $\mathcal{M}_+^f(A_{[a,b]})$  is metrizable [17]. Consider a sequence of finite measures  $\nu_n$  on  $A_{[a,b]}$ , converging weakly to a finite measure  $\nu$ . Let  $g$  be a bounded, continuous function on  $[a, b]$ , vanishing at its end-points,  $a$  and  $b$ . Then, it is clear that the function  $h_g$  defined in (25) is bounded and continuous on  $A_{[a,b]}$ . Therefore,  $\int h_g d\nu_n$  converges to  $\int h_g d\nu$ , where the integrals are over  $A_{[a,b]}$ . Hence, by (26),  $\int g dD(\nu_n)$  converges to  $\int g dD(\nu)$ . It follows that  $\nu_n$  converges to  $\nu$  in the tempered topology.

It was shown in the proof of Theorem 2.2 that  $\hat{\Phi}_n/n$  (which was denoted  $\Phi_n/n$  there!) satisfy the LDP on  $\mathcal{M}_+^f(A_{[a,b]})$  equipped with the weak topology. Since the map  $D(\cdot)$  is continuous, it follows by the contraction principle [5, Theorem 4.2.1] that  $\Psi_n/n$  satisfy the LDP on  $\mathcal{M}_+^f([a, b])$  equipped with the tempered topology, with a good rate function  $K_{[a,b]}(\cdot)$ , which can be expressed as the solution of a minimisation problem.

It remains to strengthen this LDP to the weak topology on  $\mathcal{M}_+^f([a, b])$ . We do this by showing that the sequence of random variables  $\Psi_n$ ,  $n \in \mathbb{N}$ , is exponentially tight in the weak topology. In order to show this, fix  $\alpha > 0$ , arbitrarily large. We need to find a weakly compact subset  $K$  of  $\mathcal{M}_+^f([a, b])$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Psi_n}{n} \in K^c\right) \leq -\alpha,$$

where  $K^c$  denotes the complement of  $K$ . Fix  $\gamma > 0$  sufficiently large, and take

$$K(\gamma) = \{\nu \in \mathcal{M}_+^f([a, b]) : \nu([a, b]) \leq \gamma\}.$$

Then  $K(\gamma)$  is compact in the weak topology, as noted in the proof of Proposition 3.7. Moreover,

$$\Psi_n([a, b]) = \hat{\Phi}_n(\text{cl}(A_{[a, b]} \setminus A_b)) \leq \hat{\Phi}_n(A_{[a, b]}),$$

and so,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\Psi_n}{n} \in K(\gamma)^c\right) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\hat{\Phi}_n}{n}(A_{[a, b]}) > \gamma\right).$$

By the goodness of the rate function governing the LDP of  $\hat{\Phi}_n/n$ , the last term tends to  $-\infty$  as  $\gamma$  tends to infinity. Hence, we can choose  $\gamma$  large enough to make it smaller than  $-\alpha$ , as required.

Since  $(\Psi_n/n, n \in \mathbb{N})$  satisfy the LDP on  $\mathcal{M}_+^f([a, b])$  equipped with the tempered topology, and are exponentially tight in the weak topology, it follows by [5, Corollary 4.2.6] that the LDP also holds in the weak topology, and by [5, Lemma 1.2.18] that the rate function is good.

It remains to show that  $\Psi_n/n$  satisfy Assumptions [A1] and [A3]. The proof is very similar to the corresponding part of the proof of Theorem 2.2. Translation invariance is inherited from  $\hat{\Phi}_n/n$ , and finiteness of the mean intensity is also easy to prove using the same property for  $\Phi_n/n$  and  $F$ . To prove [A3], we use the fact that  $\Psi_n([a, b])$  is dominated by  $\hat{\Phi}_n(A_{[a, b]})$ . We omit the details, which are identical to the proof of Theorem 2.2.  $\square$

**Acknowledgements** The second author learnt of the biological problem that motivated this work at a workshop at the Mathematical Biosciences Institute, Ohio, in September 2011. He would like to thank Tom Kurtz and the MBI for the invitation. We would also like to thank the anonymous referee for comments and suggestions that significantly improved the paper, including the suggestion to study the departure process from the queue.

## References

- [1] Baxter, J., Jain, N. (1988). A Comparison Principle for Large Deviations. *Proc. Amer. Math. Soc.*, 103:1235–1240.
- [2] Blom, J., De Turck, K., Mandjes, M. (2016). Functional central limit theorems for Markov-modulated infinite-server systems. *Math. Methods Oper. Res.*, 83(3):351–372.
- [3] Bogachev, V. (2007). *Measure Theory (Volume 2)*. Springer.
- [4] Chaganty, N. R. (1997). Large deviations for joint distributions and statistical applications. *Sankhyā*, Series A 59:147–166.
- [5] Dembo, A., Zeitouni, O. (1998). *Large Deviations Techniques and Applications (2nd edition)*. Springer, New York.
- [6] Dupuis, P., Ellis, R. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics. John Wiley and Sons, New York.
- [7] Florens, D., Pham, H. (1998). Large deviation probabilities in estimation of Poisson random measures. *Stochastic Process. Appl.*, 76:117–139.
- [8] Heemskerk, M., van Leeuwaarden, J., Mandjes, M. (2017). Scaling limits for infinite-server systems in a random environment. *Stoch. Syst.*, 7:1–31.
- [9] Jansen, H., Mandjes, M., De Turck, K., Wittevrongel, S. (2016). A large deviations principle for infinite-server queues in a random environment. *Queueing Syst.*, 82(1):199–235.
- [10] Léonard, C. (2000). Large deviations for Poisson random measures and processes with independent increments. *Stochastic Process. Appl.*, 85:93–121.
- [11] Lestas, I., Paulsson, J., Ross, N., Vinnicombe, G. (2008). Noise in Gene Regulatory Networks. *IEEE Trans. Automat. Control*, 53:189–200.
- [12] O’Cinneide, C., Purdue, P. (1986). The  $M/M/\infty$  queue in a Random Environment. *J. Appl. Probab.*, 23:175–184.
- [13] Paulsson, J. (2005). Models of stochastic gene expression. *Phys. Life Rev.*, 2:157–175.

- [14] Reyes, J., Chen, J., Stewart-Ornstein, J., Karhohs, K., Mock, C., Lahav, G. (2018). Fluctuations in p53 signalling allow escape from cell cycle arrest, *Molecular Cell*, 71:581–591.e5.
- [15] Stockholm, D., Edom-Vovard, F., Coutant, S., Sanatine, P., Yamagata, Y., Corre, G., Le Guillou, L., Neildez-Nguyen, T., Paldi, A. (2010). Bistable cell fate specification as a result of stochastic fluctuations and collective spatial cell behaviour. *PLoS ONE*, 5:e14441.
- [16] Schreiber, T. (2003). Large deviation principle for empirical measures generated by Cox point processes. *Colloq. Math.*, 97(1):87–106.
- [17] Varadarajan, V. (1958). Weak Convergence of Measures on Separable Metric Spaces. *Sankhyā*, 19:15–22.
- [18] Yamasaki, Y. (1975). Kolmogorov’s Extension Theorem for Infinite Measures. *Publ. RIMS, Kyoto Univ.*, 10:381–411.